Course Project

Jie Tao

9/16/2020

Contents

Prediction Assignment Writeup

Summary

The project utilizes activity data collected by wearable devices such as *Jawbone Up, Nike FuelBand,* and *Fitbit* to predict how well 6 participants perform in weight lifting exercise. The goal of this project is to use a variable "classe" in the raw training set to predict the same variable in validation set. I started by separating the raw training data into training and testing sets. Then, pre-processing the training set by removing all variable with last amount of missing values, near-zero-variance variables, and identification variables. After training three different models (decision three model, generalized boost model, and random forest model), I selected the **Random Forest Model (RFM)** as the final model based on its **accuracy** of **0.988**. Then, I applied the RFM to the validation set and find that the predicted *classe* is **B A B A A E D B A A B C B A E E A B B B**.

Data Splits and Exploration

In this section, I read the raw data into *act_train* and *act_validation*. Then, I splitted the *act_train* into training and testing sets. The overview of the two sets is shown in the end.

```
# Read raw data
act_train <- read.csv("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv")
act_validation <- read.csv("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv")

# Create training and testing sets
intrain <- createDataPartition(act_train$classe, p = 0.7, list = FALSE)
training <- act_train[intrain, ]
testing <- act_train[-intrain, ]
dim(training)
```

```
## [1] 13737   160
```

```
dim(testing)
```

```
## [1] 5885  160
```

```
# Glimpse the training data
library(skimr)
skim_without_charts(training)
```

Table 1: Data summary

Name	training
Number of rows	13737
Number of columns	160
Column type frequency:	
character	37
numeric	123
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
user_name	0	1	5	8	0	6	0
cvtd_timestamp	0	1	16	16	0	20	0
new_window	0	1	2	3	0	2	0
kurtosis_roll_belt	0	1	0	9	13453	278	0
kurtosis_picth_belt	0	1	0	9	13453	231	0
kurtosis_yaw_belt	0	1	0	7	13453	2	0
skewness_roll_belt	0	1	0	9	13453	278	0
skewness_roll_belt.1	0	1	0	9	13453	241	0
skewness_yaw_belt	0	1	0	7	13453	2	0
max_yaw_belt	0	1	0	7	13453	62	0
min_yaw_belt	0	1	0	7	13453	62	0
amplitude_yaw_belt	0	1	0	7	13453	4	0
kurtosis_roll_arm	0	1	0	8	13453	231	0
kurtosis_picth_arm	0	1	0	8	13453	229	0
kurtosis_yaw_arm	0	1	0	8	13453	278	0
skewness_roll_arm	0	1	0	8	13453	232	0
skewness_pitch_arm	0	1	0	8	13453	229	0
skewness_yaw_arm	0	1	0	8	13453	279	0
kurtosis_roll_dumbbell	0	1	0	7	13453	279	0
kurtosis_picth_dumbbell	0	1	0	7	13453	280	0
kurtosis_yaw_dumbbell	0	1	0	7	13453	2	0
skewness_roll_dumbbell	0	1	0	7	13453	281	0
skewness_pitch_dumbbell	0	1	0	7	13453	283	0
skewness_yaw_dumbbell	0	1	0	7	13453	2	0
max_yaw_dumbbell	0	1	0	7	13453	63	0
min_yaw_dumbbell	0	1	0	7	13453	63	0

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
amplitude_yaw_dumbbell	0	1	0	7	13453	3	0
kurtosis_roll_forearm	0	1	0	7	13453	226	0
kurtosis_picth_forearm	0	1	0	7	13453	226	0
kurtosis_yaw_forearm	0	1	0	7	13453	2	0
skewness_roll_forearm	0	1	0	7	13453	228	0
skewness_pitch_forearm	0	1	0	7	13453	225	0
skewness_yaw_forearm	0	1	0	7	13453	2	0
max_yaw_forearm	0	1	0	7	13453	38	0
min_yaw_forearm	0	1	0	7	13453	38	0
amplitude_yaw_forearm	0	1	0	7	13453	3	0
classe	0	1	1	1	0	5	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
X	0	1.00	9824.42	5660.96	1.00000e+00	4904.00	9831.00	14716.00	1.962100e+04
raw_timestamp_part_1	0	1.00	1322826587	286184.67	1.32249e+09	1322673100	1322832918	1323084261	1.323095e+09
raw_timestamp_part_2	0	1.00	502221.07	287801.82	2.94000e+02	252290.00	500321.00	748406.00	9.988010e+05
num_window	0	1.00	433.10	248.00	1.00000e+00	224.00	428.00	647.00	8.640000e+02
roll_belt	0	1.00	64.52	62.81	-2.86000e+01	1.10	114.00	123.00	1.620000e+02
pitch_belt	0	1.00	0.12	22.41	-5.58000e+01	1.67	5.26	14.60	6.030000e+01
yaw_belt	0	1.00	-10.74	95.61	-1.80000e+02	-88.30	-12.70	14.20	1.790000e+02
total_accel_belt	0	1.00	11.33	7.76	0.00000e+00	3.00	17.00	18.00	2.900000e+01
max_roll_belt	13453	0.02	-9.46	94.83	-9.43000e+01	-88.12	-8.70	7.10	1.800000e+02
max_picth_belt	13453	0.02	12.63	8.10	3.00000e+00	5.00	18.00	19.00	3.000000e+01
min_roll_belt	13453	0.02	-12.52	93.85	-1.80000e+02	-88.50	-13.80	3.20	1.730000e+02
min_pitch_belt	13453	0.02	10.48	7.51	0.00000e+00	3.00	16.00	17.00	2.300000e+01
amplitude_roll_belt	13453	0.02	3.06	21.39	0.00000e+00	0.30	1.00	2.00	3.600000e+02
amplitude_pitch_belt	13453	0.02	2.15	2.48	0.00000e+00	1.00	1.00	2.00	1.200000e+01
var_total_accel_belt	13453	0.02	1.00	2.44	0.00000e+00	0.10	0.20	0.30	1.650000e+01
avg_roll_belt	13453	0.02	65.56	63.61	-2.74000e+01	1.10	114.35	123.10	1.574000e+02
stddev_roll_belt	13453	0.02	1.37	2.61	0.00000e+00	0.10	0.40	0.70	1.420000e+01
var_roll_belt	13453	0.02	8.67	26.12	0.00000e+00	0.00	0.10	0.40	2.007000e+02
avg_pitch_belt	13453	0.02	0.39	22.14	-5.14000e+01	2.15	5.25	14.22	5.970000e+01
stddev_pitch_belt	13453	0.02	0.60	0.65	0.00000e+00	0.20	0.40	0.70	4.000000e+00
var_pitch_belt	13453	0.02	0.77	1.78	0.00000e+00	0.00	0.10	0.50	1.620000e+01
avg_yaw_belt	13453	0.02	-11.48	94.00	-1.38300e+02	-88.30	-12.60	3.83	1.735000e+02
stddev_yaw_belt	13453	0.02	0.96	6.49	0.00000e+00	0.10	0.30	0.60	1.092000e+02
var_yaw_belt	13453	0.02	42.96	707.77	0.00000e+00	0.01	0.09	0.36	1.192847e+04
gyros_belt_x	0	1.00	0.00	0.21	-1.04000e+00	-0.03	0.03	0.11	2.200000e+00
gyros_belt_y	0	1.00	0.04	0.08	-5.30000e-01	0.00	0.02	0.11	6.400000e-01

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
gyros_belt_z	0	1.00	-0.13	0.24	-1.46000e+00	-0.20	-0.10	-0.02	1.620000e+00
accel_belt_x	0	1.00	-5.38	29.75	-1.20000e+02	-21.00	-15.00	-5.00	7.800000e+01
accel_belt_y	0	1.00	30.13	28.51	-6.50000e+01	3.00	35.00	61.00	1.640000e+02
accel_belt_z	0	1.00	-72.82	100.60	-2.75000e+02	-162.00	-152.00	27.00	1.040000e+02
magnet_belt_x	0	1.00	56.00	64.53	-5.20000e+01	9.00	35.00	60.00	4.850000e+02
magnet_belt_y	0	1.00	593.54	35.98	3.54000e+02	581.00	601.00	610.00	6.690000e+02
magnet_belt_z	0	1.00	-345.72	64.89	-6.23000e+02	-375.00	-319.00	-306.00	2.890000e+02
roll_arm	0	1.00	17.93	72.63	-1.80000e+02	-31.80	0.00	77.30	1.800000e+02
pitch_arm	0	1.00	-4.81	30.65	-8.88000e+01	-26.10	0.00	10.90	8.850000e+01
yaw_arm	0	1.00	-0.85	71.00	-1.80000e+02	-42.80	0.00	45.20	1.800000e+02
total_accel_arm	0	1.00	25.47	10.57	1.00000e+00	17.00	27.00	33.00	6.600000e+01
var_accel_arm	13453	0.02	53.23	53.20	0.00000e+00	11.75	40.89	75.68	3.317000e+02
avg_roll_arm	13453	0.02	11.43	69.00	-1.66670e+02	-38.34	0.00	74.39	1.607800e+02
stddev_roll_arm	13453	0.02	11.00	16.04	0.00000e+00	1.45	5.55	14.92	1.614500e+02
var_roll_arm	13453	0.02	377.52	1749.50	0.00000e+00	2.10	30.83	222.53	2.606658e+04
avg_pitch_arm	13453	0.02	-4.95	27.11	-8.17700e+01	-22.32	0.00	7.96	7.566000e+01
stddev_pitch_arm	13453	0.02	10.25	9.15	0.00000e+00	1.71	8.13	16.10	4.341000e+01
var_pitch_arm	13453	0.02	188.58	281.88	0.00000e+00	2.91	66.15	259.12	1.884560e+03
avg_yaw_arm	13453	0.02	0.92	61.51	-1.73440e+02	-29.27	0.00	38.86	1.520000e+02
stddev_yaw_arm	13453	0.02	22.07	22.49	0.00000e+00	2.62	17.47	36.33	1.632600e+02
var_yaw_arm	13453	0.02	991.06	2382.81	0.00000e+00	6.88	305.09	1320.12	2.665319e+04
gyros_arm_x	0	1.00	0.05	2.00	-6.36000e+00	-1.32	0.08	1.57	4.870000e+00
gyros_arm_y	0	1.00	-0.26	0.86	-3.40000e+00	-0.80	-0.24	0.14	2.840000e+00
gyros_arm_z	0	1.00	0.27	0.55	-2.33000e+00	-0.07	0.25	0.72	2.950000e+00
accel_arm_x	0	1.00	-58.84	180.96	-4.04000e+02	-240.00	-41.00	84.00	4.370000e+02
accel_arm_y	0	1.00	31.64	109.82	-3.18000e+02	-55.00	12.00	138.00	3.080000e+02
accel_arm_z	0	1.00	-72.78	135.85	-6.36000e+02	-146.00	-47.00	23.00	2.920000e+02
magnet_arm_x	0	1.00	196.71	443.22	-5.84000e+02	-297.00	301.00	639.00	7.820000e+02
magnet_arm_y	0	1.00	153.82	202.42	-3.92000e+02	-13.00	196.00	321.00	5.830000e+02
magnet_arm_z	0	1.00	302.13	329.32	-5.97000e+02	118.00	440.00	544.00	6.940000e+02

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
max_roll_arm	13453	0.02	11.09	26.99	-7.31000e+01	-0.12	5.55	25.95	8.550000e+01
max_picth_arm	13453	0.02	33.97	69.25	-1.73000e+02	-2.10	22.70	94.12	1.800000e+02
max_yaw_arm	13453	0.02	35.65	10.09	4.00000e+00	30.00	35.00	41.00	6.000000e+01
min_roll_arm	13453	0.02	-21.11	28.82	-8.91000e+01	-41.90	-22.50	0.00	6.640000e+01
min_pitch_arm	13453	0.02	-36.01	59.55	-1.80000e+02	-73.23	-34.60	0.00	1.520000e+02
min_yaw_arm	13453	0.02	14.65	9.18	2.00000e+00	8.00	12.50	19.00	3.800000e+01
amplitude_roll_arm	13453	0.02	32.19	26.58	0.00000e+00	6.15	28.55	50.16	1.195000e+02
amplitude_pitch_arm	13453	0.02	69.98	65.59	0.00000e+00	9.90	57.30	116.85	3.590000e+02
amplitude_yaw_arm	13453	0.02	21.00	12.12	0.00000e+00	13.00	22.00	28.00	5.200000e+01
roll_dumbbell	0	1.00	24.13	69.83	-1.53710e+02	-17.77	48.34	67.91	1.535500e+02
pitch_dumbbell	0	1.00	-10.79	36.97	-1.48500e+02	-40.63	-21.01	17.47	1.494000e+02
yaw_dumbbell	0	1.00	1.53	82.44	-1.41810e+02	-77.65	-3.66	79.26	1.549500e+02
max_roll_dumbbell	13453	0.02	13.46	49.24	-7.01000e+01	-27.70	13.10	50.80	1.370000e+02
max_picth_dumbbell	13453	0.02	32.34	93.70	-1.12900e+02	-67.05	41.60	132.62	1.550000e+02
min_roll_dumbbell	13453	0.02	-41.75	33.38	-1.49600e+02	-59.32	-42.70	-27.25	7.320000e+01
min_pitch_dumbbell	13453	0.02	-35.27	73.22	-1.47000e+02	-92.55	-70.55	16.40	1.209000e+02
amplitude_roll_dumbbell	13453	0.02	55.21	56.42	0.00000e+00	15.18	33.91	78.10	2.564800e+02
amplitude_pitch_dumbbell	13453	0.02	67.61	68.02	0.00000e+00	16.67	41.72	104.49	2.735900e+02
total_accel_dumbbell	0	1.00	13.71	10.22	0.00000e+00	4.00	10.00	20.00	4.000000e+01
var_accel_dumbbell	13453	0.02	4.68	15.55	0.00000e+00	0.38	0.97	3.46	2.304300e+02
avg_roll_dumbbell	13453	0.02	24.92	61.77	-1.28960e+02	-10.67	50.09	62.85	1.259900e+02
stddev_roll_dumbbell	13453	0.02	21.87	25.97	0.00000e+00	4.65	11.87	28.05	1.237800e+02
var_roll_dumbbell	13453	0.02	1150.59	2471.89	0.00000e+00	21.66	140.80	786.71	1.532101e+04
avg_pitch_dumbbell	13453	0.02	-12.64	31.89	-7.07300e+01	-43.08	-18.70	12.56	9.393000e+01
stddev_pitch_dumbbell	13453	0.02	13.25	13.86	0.00000e+00	3.41	7.95	18.87	8.268000e+01
var_pitch_dumbbell	13453	0.02	366.87	737.06	0.00000e+00	11.61	63.15	356.09	6.836020e+03
avg_yaw_dumbbell	13453	0.02	-1.28	77.46	-1.17950e+02	-77.01	-4.85	70.73	1.349000e+02
stddev_yaw_dumbbell	13453	0.02	17.09	18.33	0.00000e+00	3.87	10.26	24.62	9.956000e+01
var_yaw_dumbbell	13453	0.02	626.81	1259.37	0.00000e+00	14.99	105.35	606.17	9.912850e+03
gyros_dumbbell_x	0	1.00	0.17	0.39	-1.99000e+00	-0.03	0.13	0.35	2.220000e+00
gyros_dumbbell_y	0	1.00	0.04	0.49	-2.10000e+00	-0.14	0.03	0.21	4.370000e+00
gyros_dumbbell_z	0	1.00	-0.15	0.32	-2.30000e+00	-0.31	-0.13	0.03	1.870000e+00
accel_dumbbell_x	0	1.00	-28.48	67.01	-2.37000e+02	-50.00	-9.00	11.00	2.350000e+02

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
accel_dumbbell_y	0	1.00	52.97	80.90	-1.89000e+02	-8.00	42.00	111.00	3.150000e+02
accel_dumbbell_z	0	1.00	-38.77	108.98	-3.19000e+02	-142.00	-1.00	37.00	3.180000e+02
magnet_dumbbell_x	0	1.00	-330.27	337.45	-6.43000e+02	-536.00	-479.00	-307.00	5.920000e+02
magnet_dumbbell_y	0	1.00	223.09	325.51	-3.60000e+03	233.00	311.00	391.00	6.330000e+02
magnet_dumbbell_z	0	1.00	46.81	140.11	-2.49000e+02	-44.00	14.00	96.00	4.520000e+02
roll_forearm	0	1.00	35.13	107.45	-1.80000e+02	0.00	22.50	141.00	1.800000e+02
pitch_forearm	0	1.00	10.70	27.81	-7.25000e+01	0.00	9.18	28.10	8.980000e+01
yaw_forearm	0	1.00	19.33	102.88	-1.80000e+02	-68.20	0.00	110.00	1.800000e+02
max_roll_forearm	13453	0.02	23.75	31.68	-6.40000e+01	0.00	26.50	45.28	8.980000e+01
max_picth_forearm	13453	0.02	80.44	95.20	-1.51000e+02	0.00	111.50	173.25	1.800000e+02
min_roll_forearm	13453	0.02	-0.48	23.25	-7.25000e+01	-5.95	0.00	12.80	6.210000e+01
min_pitch_forearm	13453	0.02	-51.87	111.68	-1.80000e+02	-174.00	-49.40	5.30	1.670000e+02
amplitude_roll_forearm	13453	0.02	24.22	26.32	0.00000e+00	1.10	17.18	37.72	1.260000e+02
amplitude_pitch_forearm	13453	0.02	132.31	144.67	0.00000e+00	1.95	83.70	346.75	3.590000e+02
total_accel_forearm	0	1.00	34.83	10.04	0.00000e+00	30.00	36.00	41.00	7.300000e+01
var_accel_forearm	13453	0.02	32.55	33.30	0.00000e+00	6.42	20.04	50.31	1.726100e+02
avg_roll_forearm	13453	0.02	29.31	78.50	-1.77230e+02	-3.57	5.68	102.06	1.771200e+02
stddev_roll_forearm	13453	0.02	41.70	59.10	0.00000e+00	0.42	6.97	84.65	1.791700e+02
var_roll_forearm	13453	0.02	5218.96	9163.50	0.00000e+00	0.18	48.66	7166.60	3.210224e+04
avg_pitch_forearm	13453	0.02	11.34	25.58	-6.55200e+01	0.00	12.08	28.28	7.209000e+01
stddev_pitch_forearm	13453	0.02	7.84	8.94	0.00000e+00	0.34	5.29	12.21	4.775000e+01
var_pitch_forearm	13453	0.02	141.08	286.16	0.00000e+00	0.11	28.01	149.06	2.279620e+03
avg_yaw_forearm	13453	0.02	19.39	79.37	-1.55060e+02	-25.45	0.00	88.86	1.692400e+02
stddev_yaw_forearm	13453	0.02	42.42	49.79	0.00000e+00	0.52	24.74	71.15	1.975100e+02
var_yaw_forearm	13453	0.02	4269.59	6984.59	0.00000e+00	0.27	612.21	5067.01	3.900933e+04
gyros_forearm_x	0	1.00	0.16	0.63	-3.36000e+00	-0.21	0.05	0.58	3.260000e+00
gyros_forearm_y	0	1.00	0.05	2.16	-7.02000e+00	-1.48	0.03	1.61	6.130000e+00
gyros_forearm_z	0	1.00	0.14	0.60	-8.09000e+00	-0.18	0.08	0.49	4.310000e+00
accel_forearm_x	0	1.00	-61.87	180.43	-4.96000e+02	-178.00	-57.00	75.00	4.770000e+02
accel_forearm_y	0	1.00	165.13	200.96	-5.85000e+02	58.00	202.00	314.00	5.890000e+02
accel_forearm_z	0	1.00	-54.73	138.29	-4.10000e+02	-181.00	-38.00	26.00	2.910000e+02

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
magnet_forearm_x	0	1.00	-314.15	345.43	-1.28000e+03	-617.00	-379.00	-79.00	6.660000e+02
magnet_forearm_y	0	1.00	383.69	509.19	-8.90000e+02	10.00	595.00	739.00	1.480000e+03
magnet_forearm_z	0	1.00	395.44	369.77	-9.73000e+02	194.00	514.00	653.00	1.090000e+03

Data Pre-processing

Given that many variables have large amount of missing values and zero variance, I excluded them from the final training and testing sets. Also, 5 identification variables such as *user_name* were left out. Eventually, there are 54 variables left in both training and testing sets.

```
# Remove all variables with missing values
library(tidyverse)
training <- training %>%
    select_if(~!any(is.na(.))) %>%
    mutate(classe = factor(classe))
testing <- testing%>%
    select_if(~!any(is.na(.))) %>%
    mutate(classe = factor(classe))

# remove all variables with near zero variance
nzv <- nearZeroVar(training)
training <- training[, -nzv]
testing <- testing[, -nzv]

# remove all id variables
training <- training[, -(1:5)]
testing <- testing[, -(1:5)]

dim(training)
```

```
## [1] 13737    54
```

```
dim(testing)
```

```
## [1] 5885   54
```

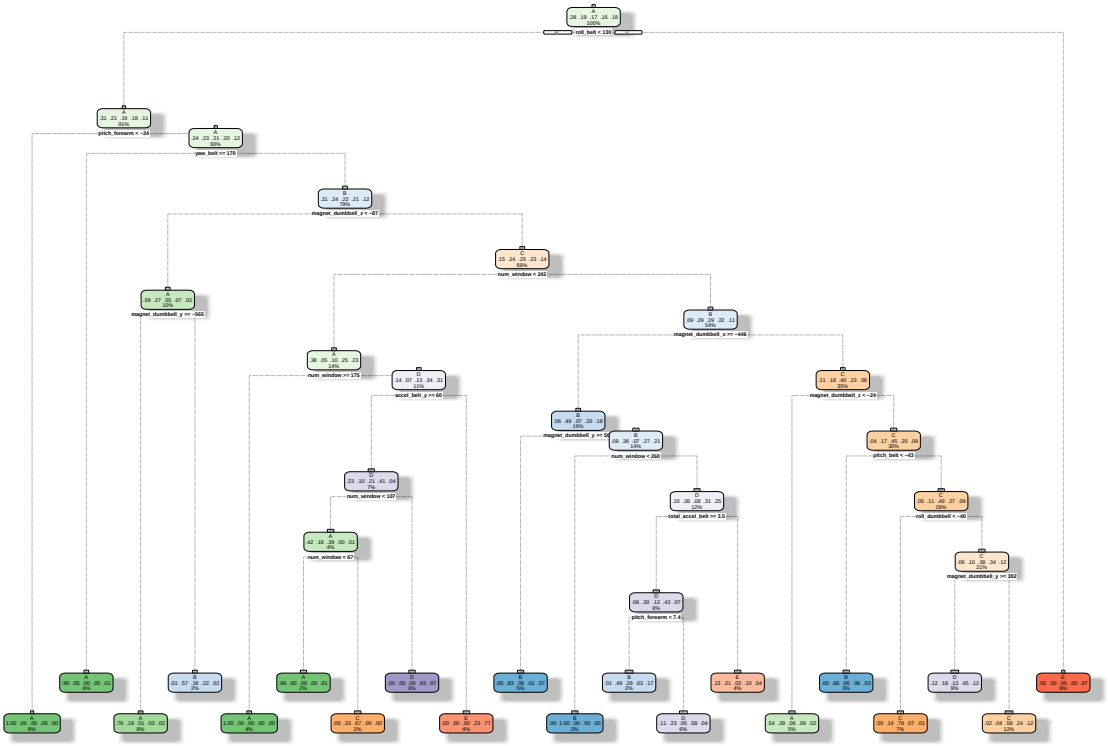
Model Training and Selection

In total, I have trained three different models: Decision Tree Model (DTM), Generalized Boost Model (GBM), and Random Forest Model (RFM). The RFM was chosen as the final model due to its highest accuracy at 0.988.

Model Training

```
set.seed(121)
modfit_rpart <- rpart(classe ~ .,
                      method = "class",
                      data = training)
fancyRpartPlot(modfit_rpart)
```

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




```
# prediction
pred_rpart <- predict(modfit_rpart, newdata = testing, type = "class")
conf_mat_rpart <-confusionMatrix(pred_rpart, testing$classe)
conf_mat_rpart
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1505  214   22   55   19
##          B    3  609   81   43   52
##          C    8  142  811  215   89
##          D   97  134  108  564  120
##          E   61   40    4   87  802
##
```

```
## Overall Statistics
##
##                Accuracy : 0.7291
##                  95% CI : (0.7176, 0.7405)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.6569
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.8990   0.5347   0.7904  0.58506   0.7412
## Specificity            0.9264   0.9623   0.9066  0.90673   0.9600
## Pos Pred Value         0.8292   0.7728   0.6411  0.55132   0.8068
## Neg Pred Value         0.9585   0.8960   0.9535  0.91773   0.9428
## Prevalence             0.2845   0.1935   0.1743  0.16381   0.1839
## Detection Rate         0.2557   0.1035   0.1378  0.09584   0.1363
## Detection Prevalence   0.3084   0.1339   0.2150  0.17383   0.1689
## Balanced Accuracy      0.9127   0.7485   0.8485  0.74589   0.8506
```

```
# Generalized boosted model
set.seed(234)
control_gbm <- trainControl(method = "repeatedcv", number = 5, repeats = 1)
modfit_bgm <- train(classe ~ .,
                    method = "gbm",
                    trControl = control_gbm,
                    verbose = FALSE,
                    data = training)
# prediction
pred_gbm <- predict(modfit_bgm, newdata = testing)
conf_mat_bgm <- confusionMatrix(pred_gbm, testing$classe)
conf_mat_bgm
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1669    8    0    0    0
##          B    5 1116    7   11    3
##          C    0   15 1015   11    1
##          D    0    0    4  941   11
##          E    0    0    0    1 1067
##
## Overall Statistics
##
##                Accuracy : 0.9869
##                  95% CI : (0.9837, 0.9897)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9834
```

```
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9970   0.9798   0.9893   0.9761   0.9861
## Specificity            0.9981   0.9945   0.9944   0.9970   0.9998
## Pos Pred Value         0.9952   0.9772   0.9741   0.9843   0.9991
## Neg Pred Value         0.9988   0.9952   0.9977   0.9953   0.9969
## Prevalence             0.2845   0.1935   0.1743   0.1638   0.1839
## Detection Rate         0.2836   0.1896   0.1725   0.1599   0.1813
## Detection Prevalence   0.2850   0.1941   0.1771   0.1624   0.1815
## Balanced Accuracy      0.9976   0.9872   0.9919   0.9865   0.9930
```

```
# Random forest
set.seed(456)
control_rf <- trainControl(method = "cv", number = 3, verboseIter = FALSE)
modfit_rf <- train(classe ~ .,
                   method = "rf",
                   trControl = control_rf,
                   data = training)
# prediction
pred_rf <- predict(modfit_rf, newdata = testing)
conf_mat_rf <- confusionMatrix(pred_rf, testing$classe)
conf_mat_rf
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1673    6    0    0    0
##          B    0 1132    1    0    0
##          C    0    1 1025    3    0
##          D    0    0    0  961    1
##          E    1    0    0    0 1081
##
## Overall Statistics
##
##                Accuracy : 0.9978
##                  95% CI : (0.9962, 0.9988)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9972
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9994   0.9939   0.9990   0.9969   0.9991
## Specificity            0.9986   0.9998   0.9992   0.9998   0.9998
## Pos Pred Value         0.9964   0.9991   0.9961   0.9990   0.9991
```

```
## Neg Pred Value         0.9998   0.9985   0.9998   0.9994   0.9998
## Prevalence             0.2845   0.1935   0.1743   0.1638   0.1839
## Detection Rate         0.2843   0.1924   0.1742   0.1633   0.1837
## Detection Prevalence   0.2853   0.1925   0.1749   0.1635   0.1839
## Balanced Accuracy      0.9990   0.9968   0.9991   0.9983   0.9994
```

Final Model Selection

```
# Final model selection
which.max(c(conf_mat_rpart$overall["Accuracy"], conf_mat_bgm$overall["Accuracy"], conf_mat_rf$overall["
```

```
## Accuracy
##        3
```

Predication

The results from RFM were applied to the validation set to predict the results of *classe*. Based on the prediction, *classe* follows the sequence of **B A B A A E D B A A B C B A E E A B B B**

```
pred_val <- predict(modfit_rf, newdata = act_validation)
pred_val
```

```
##  [1] 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
```