Coursera - Practical Machine Learning

By Mandy Jiang (04/04/2022)

Background and Description

The goal of this project is to use data from accelerometers on the belt, forearm, arm, and dumbell of 6 participants, to predict the manner in which they did the exercise. They were asked to perform barbell lifts correctly and incorrectly in 5 different ways. Variable "classe" in the training set shows which way they did in exercise, in which Class A corresponds to the specified execution of the exercise, while the other 4 classes correspond to common mistakes. (The Weight Lifting Exercise Datasethttp://web.archive.org/web/20161224072740/http:/groupware.les.inf.puc-rio.br/har)

In [38]:
```
library(ggplot2)
library(caret)
library(lattice)
library(e1071)
library(rpart)
library(kernlab)
library(randomForest)
library(gbm)
```

Data preprocessing

We are going to train and test the model on the training set only, and leave the testing set for final validation. Therefore, the data cleaning and preprocessin will be applied on training set only.

In [13]:
```
train_file = 'https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv'
test_file = 'https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv'
train = read.csv(train_file, sep=',', header=TRUE)
dim(train)
test = read.csv(test_file, sep=',', header=TRUE)
dim(test)
```

1. 19622
2. 160

1. 20
2. 160

We got 19622 observations and 160 variables in training dataset, while we got 20 observations and 160 variables in testing dataset. Next, we are going to remove variables with missing values and those not relevant to the prediction of classes in the training set.

In [14]:
```
train = train[, colSums(is.na(train)) == 0]
train = train[,-c(1:7)]
dim(train)
```

1. 19622
2. 86

Then we are going to remove variables with almost zero variance across observations.

In [15]:
```
nvz = nearZeroVar(train)
train = train[,-nvz]
dim(train)
```

1. 19622
2. 53

Data splitting

In [22]:
```
set.seed(1886)
inTrain = createDataPartition(y=train$classe, p=0.7, list=FALSE)
training = train[inTrain,]
validation = train[-inTrain,]
rbind("original dataset" = dim(train),"training set" = dim(training), "validation set" = dim(validation))
```

original dataset	19622	53
training set	13737	53
validation set	5885	53

Prediction model setup

We will use decision trees, random forests, Gradient Boosted Trees, and Support Vector Machine to predict the outcomes. We will also select the best performance model and look at the predictions on the testing dataset.

Set up control for training with 5-fold cross validation.

In [23]:
```
control = trainControl(method="cv", number=5, verboseIter=FALSE)
```

Decision tree

In [30]:
```
modFit1 = train(classe~., data=training, method="rpart", trControl = control)
pred1 = predict(modFit1, validation)
cm1 = confusionMatrix(pred1, factor(validation$classe))
cm1
```

```
Confusion Matrix and Statistics

          Reference
Prediction    A    B    C    D    E
         A 1492  494  493  437  145
         B   26  361   29  177  144
         C  123  284  504  350  264
         D    0    0    0    0    0
         E   33    0    0    0  529

Overall Statistics

               Accuracy : 0.4904
                 95% CI : (0.4775, 0.5033)
    No Information Rate : 0.2845
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.3339

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: A Class: B Class: C Class: D Class: E
Sensitivity            0.8913  0.31694  0.49123   0.0000  0.48891
Specificity            0.6274  0.92078  0.78987   1.0000  0.99313
Pos Pred Value         0.4874  0.48982  0.33049      NaN  0.94128
Neg Pred Value         0.9356  0.84887  0.88028   0.8362  0.89611
Prevalence             0.2845  0.19354  0.17434   0.1638  0.18386
Detection Rate         0.2535  0.06134  0.08564   0.0000  0.08989
Detection Prevalence   0.5201  0.12523  0.25913   0.0000  0.09550
Balanced Accuracy      0.7593  0.61886  0.64055   0.5000  0.74102
```

Random Forests

In [31]:
```
modFit2 = train(classe~., data=training, method="rf", trControl = control)
pred2 = predict(modFit2, validation)
cm2 = confusionMatrix(pred2, factor(validation$classe))
cm2
```

```
Confusion Matrix and Statistics

          Reference
Prediction    A    B    C    D    E
         A 1671    9    0    0    0
         B    2 1129    4    0    1
         C    0    1 1019    8    0
         D    0    0    3  956    3
         E    1    0    0    0 1078

Overall Statistics

               Accuracy : 0.9946
                 95% CI : (0.9923, 0.9963)
    No Information Rate : 0.2845
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9931

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: A Class: B Class: C Class: D Class: E
Sensitivity            0.9982   0.9912   0.9932   0.9917   0.9963
Specificity            0.9979   0.9985   0.9981   0.9988   0.9998
Pos Pred Value         0.9946   0.9938   0.9912   0.9938   0.9991
Neg Pred Value         0.9993   0.9979   0.9986   0.9984   0.9992
Prevalence             0.2845   0.1935   0.1743   0.1638   0.1839
Detection Rate         0.2839   0.1918   0.1732   0.1624   0.1832
Detection Prevalence   0.2855   0.1930   0.1747   0.1635   0.1833
Balanced Accuracy      0.9980   0.9949   0.9957   0.9952   0.9980
```

Gradient Boosted Trees

In [35]:
```
modFit3 = train(classe~., data=training, method="gbm", trControl = control)
pred3 = predict(modFit3, validation)
cm3 = confusionMatrix(pred3, factor(validation$classe))
cm3
```

```
Iter   TrainDeviance   ValidDeviance   StepSize   Improve
     1        1.6094             nan     0.1000    0.1263
     2        1.5253             nan     0.1000    0.0890
     3        1.4670             nan     0.1000    0.0674
     4        1.4223             nan     0.1000    0.0547
     5        1.3865             nan     0.1000    0.0490
     6        1.3545             nan     0.1000    0.0445
     7        1.3263             nan     0.1000    0.0376
     8        1.3030             nan     0.1000    0.0324
     9        1.2818             nan     0.1000    0.0361
    10        1.2583             nan     0.1000    0.0269
    20        1.1047             nan     0.1000    0.0170
    40        0.9324             nan     0.1000    0.0080
    60        0.8260             nan     0.1000    0.0070
    80        0.7464             nan     0.1000    0.0034
   100        0.6847             nan     0.1000    0.0033
   120        0.6327             nan     0.1000    0.0029
   140        0.5881             nan     0.1000    0.0029
   150        0.5684             nan     0.1000    0.0022

Iter   TrainDeviance   ValidDeviance   StepSize   Improve
     1        1.6094             nan     0.1000    0.1844
     2        1.4906             nan     0.1000    0.1284
     3        1.4077             nan     0.1000    0.1009
     4        1.3424             nan     0.1000    0.0840
     5        1.2888             nan     0.1000    0.0732
     6        1.2422             nan     0.1000    0.0717
     7        1.1979             nan     0.1000    0.0582
     8        1.1604             nan     0.1000    0.0494
     9        1.1287             nan     0.1000    0.0516
    10        1.0959             nan     0.1000    0.0431
    20        0.8944             nan     0.1000    0.0190
    40        0.6803             nan     0.1000    0.0092
    60        0.5537             nan     0.1000    0.0084
    80        0.4653             nan     0.1000    0.0051
   100        0.3984             nan     0.1000    0.0032
   120        0.3473             nan     0.1000    0.0036
   140        0.3050             nan     0.1000    0.0026
   150        0.2888             nan     0.1000    0.0019

Iter   TrainDeviance   ValidDeviance   StepSize   Improve
     1        1.6094             nan     0.1000    0.2312
     2        1.4626             nan     0.1000    0.1615
     3        1.3611             nan     0.1000    0.1242
     4        1.2813             nan     0.1000    0.0981
     5        1.2187             nan     0.1000    0.0965
     6        1.1586             nan     0.1000    0.0799
     7        1.1079             nan     0.1000    0.0697
     8        1.0651             nan     0.1000    0.0569
     9        1.0288             nan     0.1000    0.0708
    10        0.9849             nan     0.1000    0.0571
    20        0.7512             nan     0.1000    0.0265
    40        0.5294             nan     0.1000    0.0114
    60        0.4062             nan     0.1000    0.0081
    80        0.3225             nan     0.1000    0.0057
   100        0.2640             nan     0.1000    0.0036
   120        0.2210             nan     0.1000    0.0020
   140        0.1872             nan     0.1000    0.0018
   150        0.1740             nan     0.1000    0.0012

Iter   TrainDeviance   ValidDeviance   StepSize   Improve
     1        1.6094             nan     0.1000    0.1329
     2        1.5236             nan     0.1000    0.0872
     3        1.4653             nan     0.1000    0.0661
     4        1.4208             nan     0.1000    0.0521
     5        1.3857             nan     0.1000    0.0515
     6        1.3524             nan     0.1000    0.0380
     7        1.3278             nan     0.1000    0.0399
     8        1.3025             nan     0.1000    0.0338
     9        1.2804             nan     0.1000    0.0330
    10        1.2585             nan     0.1000    0.0301
    20        1.1056             nan     0.1000    0.0176
    40        0.9317             nan     0.1000    0.0081
    60        0.8248             nan     0.1000    0.0069
    80        0.7447             nan     0.1000    0.0039
   100        0.6799             nan     0.1000    0.0028
   120        0.6294             nan     0.1000    0.0032
   140        0.5839             nan     0.1000    0.0029
   150        0.5649             nan     0.1000    0.0025

Iter   TrainDeviance   ValidDeviance   StepSize   Improve
     1        1.6094             nan     0.1000    0.1851
     2        1.4895             nan     0.1000    0.1346
     3        1.4046             nan     0.1000    0.0996
     4        1.3403             nan     0.1000    0.0842
     5        1.2858             nan     0.1000    0.0757
     6        1.2365             nan     0.1000    0.0647
     7        1.1949             nan     0.1000    0.0605
     8        1.1562             nan     0.1000    0.0519
     9        1.1224             nan     0.1000    0.0447
    10        1.0937             nan     0.1000    0.0399
    20        0.8921             nan     0.1000    0.0228
    40        0.6816             nan     0.1000    0.0137
    60        0.5545             nan     0.1000    0.0076
    80        0.4641             nan     0.1000    0.0025
   100        0.3979             nan     0.1000    0.0037
   120        0.3469             nan     0.1000    0.0024
   140        0.3059             nan     0.1000    0.0021
   150        0.2886             nan     0.1000    0.0022

Iter   TrainDeviance   ValidDeviance   StepSize   Improve
     1        1.6094             nan     0.1000    0.2278
     2        1.4630             nan     0.1000    0.1608
     3        1.3604             nan     0.1000    0.1223
     4        1.2827             nan     0.1000    0.1065
     5        1.2153             nan     0.1000    0.0900
     6        1.1597             nan     0.1000    0.0800
     7        1.1096             nan     0.1000    0.0662
     8        1.0673             nan     0.1000    0.0687
     9        1.0242             nan     0.1000    0.0586
    10        0.9852             nan     0.1000    0.0494
    20        0.7546             nan     0.1000    0.0209
    40        0.5258             nan     0.1000    0.0107
    60        0.4043             nan     0.1000    0.0066
    80        0.3238             nan     0.1000    0.0034
   100        0.2662             nan     0.1000    0.0025
   120        0.2240             nan     0.1000    0.0028
   140        0.1888             nan     0.1000    0.0013
   150        0.1741             nan     0.1000    0.0012

Iter   TrainDeviance   ValidDeviance   StepSize   Improve
     1        1.6094             nan     0.1000    0.1282
     2        1.5236             nan     0.1000    0.0846
     3        1.4663             nan     0.1000    0.0689
     4        1.4214             nan     0.1000    0.0539
     5        1.3862             nan     0.1000    0.0495
     6        1.3538             nan     0.1000    0.0366
     7        1.3290             nan     0.1000    0.0405
     8        1.3032             nan     0.1000    0.0343
     9        1.2815             nan     0.1000    0.0309
    10        1.2607             nan     0.1000    0.0297
    20        1.1071             nan     0.1000    0.0166
    40        0.9348             nan     0.1000    0.0087
    60        0.8270             nan     0.1000    0.0062
    80        0.7445             nan     0.1000    0.0062
   100        0.6810             nan     0.1000    0.0033
   120        0.6300             nan     0.1000    0.0031
   140        0.5851             nan     0.1000    0.0023
   150        0.5642             nan     0.1000    0.0015

Iter   TrainDeviance   ValidDeviance   StepSize   Improve
     1        1.6094             nan     0.1000    0.1861
     2        1.4895             nan     0.1000    0.1273
     3        1.4054             nan     0.1000    0.1043
     4        1.3393             nan     0.1000    0.0817
     5        1.2864             nan     0.1000    0.0708
     6        1.2404             nan     0.1000    0.0716
     7        1.1958             nan     0.1000    0.0599
     8        1.1579             nan     0.1000    0.0473
     9        1.1270             nan     0.1000    0.0480
    10        1.0969             nan     0.1000    0.0406
    20        0.8892             nan     0.1000    0.0214
    40        0.6792             nan     0.1000    0.0080
    60        0.5534             nan     0.1000    0.0070
    80        0.4647             nan     0.1000    0.0046
   100        0.3985             nan     0.1000    0.0025
   120        0.3456             nan     0.1000    0.0038
   140        0.3033             nan     0.1000    0.0020
   150        0.2857             nan     0.1000    0.0029

Iter   TrainDeviance   ValidDeviance   StepSize   Improve
     1        1.6094             nan     0.1000    0.2257
     2        1.4643             nan     0.1000    0.1669
     3        1.3615             nan     0.1000    0.1208
     4        1.2837             nan     0.1000    0.0975
     5        1.2209             nan     0.1000    0.0942
     6        1.1616             nan     0.1000    0.0790
     7        1.1103             nan     0.1000    0.0750
     8        1.0643             nan     0.1000    0.0624
     9        1.0259             nan     0.1000    0.0597
    10        0.9878             nan     0.1000    0.0476
    20        0.7564             nan     0.1000    0.0238
    40        0.5313             nan     0.1000    0.0083
    60        0.4060             nan     0.1000    0.0095
    80        0.3254             nan     0.1000    0.0031
   100        0.2631             nan     0.1000    0.0030
   120        0.2193             nan     0.1000    0.0027
   140        0.1854             nan     0.1000    0.0015
   150        0.1716             nan     0.1000    0.0012

Iter   TrainDeviance   ValidDeviance   StepSize   Improve
     1        1.6094             nan     0.1000    0.1297
     2        1.5238             nan     0.1000    0.0866
     3        1.4655             nan     0.1000    0.0655
     4        1.4213             nan     0.1000    0.0545
     5        1.3854             nan     0.1000    0.0507
     6        1.3534             nan     0.1000    0.0396
     7        1.3276             nan     0.1000    0.0394
     8        1.3023             nan     0.1000    0.0345
     9        1.2799             nan     0.1000    0.0328
    10        1.2582             nan     0.1000    0.0303
    20        1.1039             nan     0.1000    0.0162
    40        0.9320             nan     0.1000    0.0090
    60        0.8224             nan     0.1000    0.0065
    80        0.7427             nan     0.1000    0.0055
   100        0.6798             nan     0.1000    0.0032
   120        0.6286             nan     0.1000    0.0028
   140        0.5841             nan     0.1000    0.0017
   150        0.5658             nan     0.1000    0.0030

Iter   TrainDeviance   ValidDeviance   StepSize   Improve
     1        1.6094             nan     0.1000    0.1878
     2        1.4877             nan     0.1000    0.1240
     3        1.4064             nan     0.1000    0.1098
     4        1.3369             nan     0.1000    0.0825
     5        1.2837             nan     0.1000    0.0722
     6        1.2376             nan     0.1000    0.0717
     7        1.1927             nan     0.1000    0.0639
     8        1.1531             nan     0.1000    0.0567
     9        1.1176             nan     0.1000    0.0424
    10        1.0899             nan     0.1000    0.0383
    20        0.8897             nan     0.1000    0.0263
    40        0.6716             nan     0.1000    0.0115
    60        0.5471             nan     0.1000    0.0077
    80        0.4598             nan     0.1000    0.0065
   100        0.3955             nan     0.1000    0.0050
   120        0.3459             nan     0.1000    0.0024
   140        0.3050             nan     0.1000    0.0022
   150        0.2849             nan     0.1000    0.0017

Iter   TrainDeviance   ValidDeviance   StepSize   Improve
     1        1.6094             nan     0.1000    0.2378
     2        1.4584             nan     0.1000    0.1581
     3        1.3583             nan     0.1000    0.1288
     4        1.2776             nan     0.1000    0.1028
     5        1.2122             nan     0.1000    0.0897
     6        1.1559             nan     0.1000    0.0698
     7        1.1109             nan     0.1000    0.0824
     8        1.0603             nan     0.1000    0.0648
     9        1.0192             nan     0.1000    0.0513
    10        0.9852             nan     0.1000    0.0552
    20        0.7476             nan     0.1000    0.0235
    40        0.5259             nan     0.1000    0.0120
    60        0.4037             nan     0.1000    0.0079
    80        0.3226             nan     0.1000    0.0030
   100        0.2658             nan     0.1000    0.0046
   120        0.2214             nan     0.1000    0.0012
   140        0.1891             nan     0.1000    0.0008
   150        0.1761             nan     0.1000    0.0016

Iter   TrainDeviance   ValidDeviance   StepSize   Improve
     1        1.6094             nan     0.1000    0.1274
     2        1.5238             nan     0.1000    0.0903
     3        1.4649             nan     0.1000    0.0631
     4        1.4213             nan     0.1000    0.0525
     5        1.3855             nan     0.1000    0.0509
     6        1.3526             nan     0.1000    0.0401
     7        1.3257             nan     0.1000    0.0370
     8        1.3021             nan     0.1000    0.0325
     9        1.2816             nan     0.1000    0.0323
    10        1.2592             nan     0.1000    0.0310
    20        1.1044             nan     0.1000    0.0185
    40        0.9327             nan     0.1000    0.0073
    60        0.8243             nan     0.1000    0.0090
    80        0.7431             nan     0.1000    0.0040
   100        0.6799             nan     0.1000    0.0025
   120        0.6277             nan     0.1000    0.0027
   140        0.5834             nan     0.1000    0.0029
   150        0.5652             nan     0.1000    0.0016

Iter   TrainDeviance   ValidDeviance   StepSize   Improve
     1        1.6094             nan     0.1000    0.1854
     2        1.4884             nan     0.1000    0.1295
     3        1.4056             nan     0.1000    0.1014
     4        1.3411             nan     0.1000    0.0779
     5        1.2901             nan     0.1000    0.0725
     6        1.2433             nan     0.1000    0.0683
     7        1.1995             nan     0.1000    0.0585
     8        1.1620             nan     0.1000    0.0592
     9        1.1254             nan     0.1000    0.0511
    10        1.0931             nan     0.1000    0.0416
    20        0.8930             nan     0.1000    0.0257
    40        0.6848             nan     0.1000    0.0135
    60        0.5544             nan     0.1000    0.0048
    80        0.4612             nan     0.1000    0.0040
   100        0.3975             nan     0.1000    0.0038
   120        0.3433             nan     0.1000    0.0028
   140        0.3048             nan     0.1000    0.0040
   150        0.2860             nan     0.1000    0.0011

Iter   TrainDeviance   ValidDeviance   StepSize   Improve
     1        1.6094             nan     0.1000    0.2380
     2        1.4577             nan     0.1000    0.1599
     3        1.3546             nan     0.1000    0.1309
     4        1.2749             nan     0.1000    0.0988
     5        1.2130             nan     0.1000    0.0926
     6        1.1539             nan     0.1000    0.0741
     7        1.1073             nan     0.1000    0.0693
     8        1.0636             nan     0.1000    0.0612
     9        1.0238             nan     0.1000    0.0635
    10        0.9841             nan     0.1000    0.0553
    20        0.7563             nan     0.1000    0.0293
    40        0.5279             nan     0.1000    0.0109
    60        0.4045             nan     0.1000    0.0080
    80        0.3179             nan     0.1000    0.0022
   100        0.2636             nan     0.1000    0.0032
   120        0.2198             nan     0.1000    0.0023
   140        0.1859             nan     0.1000    0.0017
   150        0.1737             nan     0.1000    0.0009

Iter   TrainDeviance   ValidDeviance   StepSize   Improve
     1        1.6094             nan     0.1000    0.2264
     2        1.4645             nan     0.1000    0.1608
     3        1.3622             nan     0.1000    0.1233
     4        1.2841             nan     0.1000    0.1032
     5        1.2187             nan     0.1000    0.0954
     6        1.1601             nan     0.1000    0.0812
     7        1.1099             nan     0.1000    0.0744
     8        1.0618             nan     0.1000    0.0587
     9        1.0251             nan     0.1000    0.0563
    10        0.9898             nan     0.1000    0.0547
    20        0.7591             nan     0.1000    0.0287
    40        0.5298             nan     0.1000    0.0139
    60        0.4041             nan     0.1000    0.0076
    80        0.3219             nan     0.1000    0.0042
   100        0.2664             nan     0.1000    0.0027
   120        0.2258             nan     0.1000    0.0036
   140        0.1914             nan     0.1000    0.0010
   150        0.1772             nan     0.1000    0.0024

Confusion Matrix and Statistics

          Reference
Prediction    A    B    C    D    E
         A 1641   40    0    1    2
         B   27 1070   22    3   12
         C    3   26  985   31    7
         D    2    1   18  924   13
         E    1    2    1    5 1048

Overall Statistics

               Accuracy : 0.9631
                 95% CI : (0.958, 0.9678)
    No Information Rate : 0.2845
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9534

 Mcnemar's Test P-Value : 0.003518

Statistics by Class:

                     Class: A Class: B Class: C Class: D Class: E
Sensitivity            0.9803   0.9394   0.9600   0.9585   0.9686
Specificity            0.9898   0.9865   0.9862   0.9931   0.9981
Pos Pred Value         0.9745   0.9436   0.9363   0.9645   0.9915
Neg Pred Value         0.9921   0.9855   0.9915   0.9919   0.9930
Prevalence             0.2845   0.1935   0.1743   0.1638   0.1839
Detection Rate         0.2788   0.1818   0.1674   0.1570   0.1781
Detection Prevalence   0.2862   0.1927   0.1788   0.1628   0.1796
Balanced Accuracy      0.9850   0.9630   0.9731   0.9758   0.9834
```

Support Vector Machine

In [39]:
```
modFit4 = train(classe~., data=training, method="svmLinear", trControl = control)
pred4 = predict(modFit4, validation)
cm4 = confusionMatrix(pred4, factor(validation$classe))
cm4
```

```
Confusion Matrix and Statistics

          Reference
Prediction    A    B    C    D    E
         A 1534  163   91   64   60
         B   35  803   84   40  125
         C   41   66  793  113   61
         D   52   21   25  695   61
         E   12   86   33   52  775

Overall Statistics

               Accuracy : 0.7816
                 95% CI : (0.7709, 0.7921)
    No Information Rate : 0.2845
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7223

 Mcnemar's Test P-Value : < 2.2e-16

Statistics by Class:

                     Class: A Class: B Class: C Class: D Class: E
Sensitivity            0.9164   0.7050   0.7729   0.7210   0.7163
Specificity            0.9102   0.9402   0.9422   0.9677   0.9619
Pos Pred Value         0.8023   0.7387   0.7384   0.8138   0.8090
Neg Pred Value         0.9648   0.9300   0.9516   0.9465   0.9377
Prevalence             0.2845   0.1935   0.1743   0.1638   0.1839
Detection Rate         0.2607   0.1364   0.1347   0.1181   0.1317
Detection Prevalence   0.3249   0.1847   0.1825   0.1451   0.1628
Balanced Accuracy      0.9133   0.8226   0.8575   0.8443   0.8391
```

Model evaluation

In [52]:
```
AccuracyResults = data.frame(Model = c('Decision Tree','Random Forests','Gradient Boosted Trees',
                                       'Support Vector Machine'),
                             Accuracy = rbind(cm1$overall[1],cm2$overall[1],cm3$overall[1],cm4$overall[1])
                            )
print(AccuracyResults)
```

```
                   Model  Accuracy
1          Decision Tree 0.4903993
2         Random Forests 0.9945624
3 Gradient Boosted Trees 0.9631266
4 Support Vector Machine 0.7816483
```

The best model is the Random Forest model, with 0.9945624 accuracy.

Prediction on test set

In [54]:
```
pred = predict(modFit2, newdata = test)
ValidationPredictionResults <- data.frame(problem_id=test$problem_id, predicted=pred)
print(ValidationPredictionResults)
```

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