Prediction Assignment Writeup

Fabio Turetto Rodrigues

2025-03-21

Background

This project aims to predict the manner in which participants performed exercises using data from accelerometers on their belt, forearm, arm, and dumbbell. The target variable is `classe`, which categorizes the type of movement performed.

Load Packages

```
library(caret)
```

```
## Carregando pacotes exigidos: ggplot2
```

```
## Carregando pacotes exigidos: lattice
```

```
library(randomForest)
```

```
## randomForest 4.7-1.2
```

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

```
##
## Anexando pacote: 'randomForest'
```

```
## O seguinte objeto é mascarado por 'package:ggplot2':
##
##     margin
```

```
library(gbm)
```

```
## Loaded gbm 2.2.2
```

```
## This version of gbm is no longer under development. Consider transitioning to gbm3, https://github.co
```

```
library(knitr)
```

Load Data

```r
training_url <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv"
testing_url <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv"

training <- read.csv(training_url, na.strings = c("NA", "", "#DIV/0!"))
testing <- read.csv(testing_url, na.strings = c("NA", "", "#DIV/0!"))
```

Data Preprocessing

```r
# Removing columns with too many NAs
training <- training[, colSums(is.na(training)) == 0]
testing <- testing[, colSums(is.na(testing)) == 0]

# Removing non-relevant columns (like IDs)
training <- training[, -c(1:7)]
testing <- testing[, -c(1:7)]

# Splitting into training and validation sets
set.seed(1234)
inTrain <- createDataPartition(training$classe, p = 0.75, list = FALSE)
trainSet <- training[inTrain, ]
validSet <- training[-inTrain, ]

# Convertendo a variável 'classe' para fator
trainSet$classe <- as.factor(trainSet$classe)
validSet$classe <- as.factor(validSet$classe)
```

Model Training

Random Forest Model

```r
set.seed(1234)
rf_model <- randomForest(classe ~ ., data = trainSet, importance = TRUE)
pred_rf <- predict(rf_model, validSet)
rf_accuracy <- confusionMatrix(pred_rf, validSet$classe)
rf_accuracy$overall["Accuracy"]
```

```
##  Accuracy
## 0.9959217
```

Optimized Boosted Trees Model (GBM)

```r
set.seed(1234)
gbm_model <- train(classe ~ ., data = trainSet, method = "gbm",
                   trControl = trainControl(method = "cv", number = 3),
                   verbose = FALSE,
                   tuneGrid = expand.grid(interaction.depth = 3,
```

```
                                          n.trees = 100,
                                          shrinkage = 0.1,
                                          n.minobsinnode = 10))
pred_gbm <- predict(gbm_model, validSet)
gbm_accuracy <- confusionMatrix(pred_gbm, validSet$classe)
gbm_accuracy$overall["Accuracy"]
```

```
## Accuracy
## 0.949633
```

Model Evaluation

```
kable(data.frame(
  Model = c("Random Forest", "GBM"),
  Accuracy = c(rf_accuracy$overall["Accuracy"], gbm_accuracy$overall["Accuracy"])
))
```

Model	Accuracy
Random Forest	0.9959217
GBM	0.9496330

Predictions on Test Data

```
pred_test <- predict(rf_model, testing)
write.csv(pred_test, file = "predictions.csv", row.names = FALSE)
```

Conclusion

This report summarizes the approach used for training predictive models to classify movement types. The Random Forest model performed better and was used for final predictions.