Enlik Askerbay

Report: Predicting Exercise Performance Using Accelerometer Data

Introduction

The goal of this project is to predict the manner in which participants performed barbell lifts using data collected from accelerometers on various body parts. This report details the steps taken to build and evaluate a predictive model using machine learning techniques.

Data Collection

The training and test datasets were obtained from the Human Activity Recognition Using Smartphones Dataset (HAR) repository. The training dataset consists of 19,622 observations and 160 variables, while the test dataset contains 20 observations. The target variable, "classe," represents the manner in which participants performed the exercise.

```
training_data <-
read.csv("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv")

testing_data <-
read.csv("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv")
```

Data Exploration

Initial exploration of the training dataset revealed several variables, including accelerometer readings from the belt, forearm, arm, and dumbell. Summary statistics and data distributions were examined to gain insights into the data's characteristics.

```
str(training_data)

summary(training_data)
```

Data Preprocessing

Preprocessing steps included removing irrelevant columns, handling missing values, and converting categorical variables to factors. The training dataset was split into training and testing subsets for model development and evaluation.

```
cols_to_remove <- c("X", "user_name", "raw_timestamp_part_1",
"raw_timestamp_part_2", "cvtd_timestamp")

training_data <- training_data[, !names(training_data) %in% cols_to_remove]
```

Handle missing values

training_data <- na.omit(training_data)

Convert categorical variables to factors

training_data <- lapply(training_data, function(x) if(is.factor(x)) as.factor(x) else x)

Model Building

A random forest model was chosen for its ability to handle high-dimensional data and nonlinear relationships. The caret and randomForest packages were used to build the model on the training data.

library(caret)

set.seed(123)

trainIndex <- createDataPartition(training_data$classe, p = .7, list = FALSE)

training <- training_data[trainIndex,]

testing <- training_data[-trainIndex,]

Build a random forest model

library(randomForest)

rf_model <- randomForest(classe ~ ., data = training)

Cross-Validation

Cross-validation was employed to assess the model's performance and prevent overfitting. A 5-fold cross-validation strategy was implemented to evaluate the model's accuracy and generalization ability.

Evaluate model performance using cross-validation

cv_model <- train(classe ~ ., data = training, method = "rf", trControl = trainControl(method = "cv", number = 5))

Model Evaluation

The trained model was evaluated on the testing subset using confusion matrix analysis. Performance metrics such as accuracy, precision, recall, and F1-score were calculated to assess the model's predictive power.

```
# Assess model performance
predictions <- predict(cv_model, newdata = testing)
confusionMatrix(predictions, testing$classe)
```

Prediction

The trained model was applied to the test dataset to predict the manner in which participants performed the exercise. The predictions were compared to the actual outcomes to measure the model's effectiveness.

```
# Use the trained model to predict for the test cases
test_predictions <- predict(cv_model, newdata = testing_data)
```

Conclusion

In conclusion, the predictive model achieved satisfactory performance in classifying exercise performance based on accelerometer data. Further optimization and refinement may be possible to improve the model's accuracy and robustness. Overall, this project demonstrates the potential of machine learning techniques in predicting human activity patterns using wearable sensor data.