Project3 Rmarkdown

Faith Adeyemi Allen

2024-12-13

Project Overview

Title: Predicting Weight Based on Height and Gender

In this project, we will develop a predictive model to estimate weight using height and gender as predic

Using the "height and weiht" data from project 1 and 2, we will follow the following steps to achive our

Step 1: Load Libraries

```
library(tidyverse)
library(caret)
```

Step 2: Load the Data

```
dataset <- read.csv("C:/Users/HP/OneDrive/Documents/Portfolio/Project3_PredictiveModel/data/cleaned_heig
head(dataset)
```

```
##    Gender Height_in_inchies Weight_in_pound
## 1   Male                64             128
## 2 Female                66             124
## 3 Female                62             136
## 4   Male                70             153
## 5   Male                68             144
## 6 Female                64             116
```

Step 3: Split the Data

```
set.seed(123) # For reproducibility
training_index <- createDataPartition(dataset$Weight_in_pound, p = 0.8, list = FALSE)
training_data <- dataset[training_index, ]
testing_data <- dataset[-training_index, ]
```

Step 4: Train the Model

```
model <- lm(Weight_in_pound ~ Height_in_inchies + Gender, data = training_data)
summary(model)
```

```
##
## Call:
## lm(formula = Weight_in_pound ~ Height_in_inchies + Gender, data = training_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.444  -5.889  -2.500   6.389  14.222
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          37.389    116.589   0.321    0.759
## Height_in_inchies     1.361      1.820   0.748    0.483
## GenderMale           13.944      8.767   1.591    0.163
##
## Residual standard error: 9.768 on 6 degrees of freedom
## Multiple R-squared:  0.5822, Adjusted R-squared:  0.4429
## F-statistic:  4.18 on 2 and 6 DF,  p-value: 0.07295
```

Step 5: Evaluate the Model

```
predictions <- predict(model, newdata = testing_data)
mse <- mean((predictions - testing_data$Weight_in_pound)^2)
r_squared <- 1 - (sum((predictions - testing_data$Weight_in_pound)^2) /
                  sum((testing_data$Weight_in_pound - mean(testing_data$Weight_in_pound))^2))

cat("Mean Squared Error:", mse, "\\n")
```
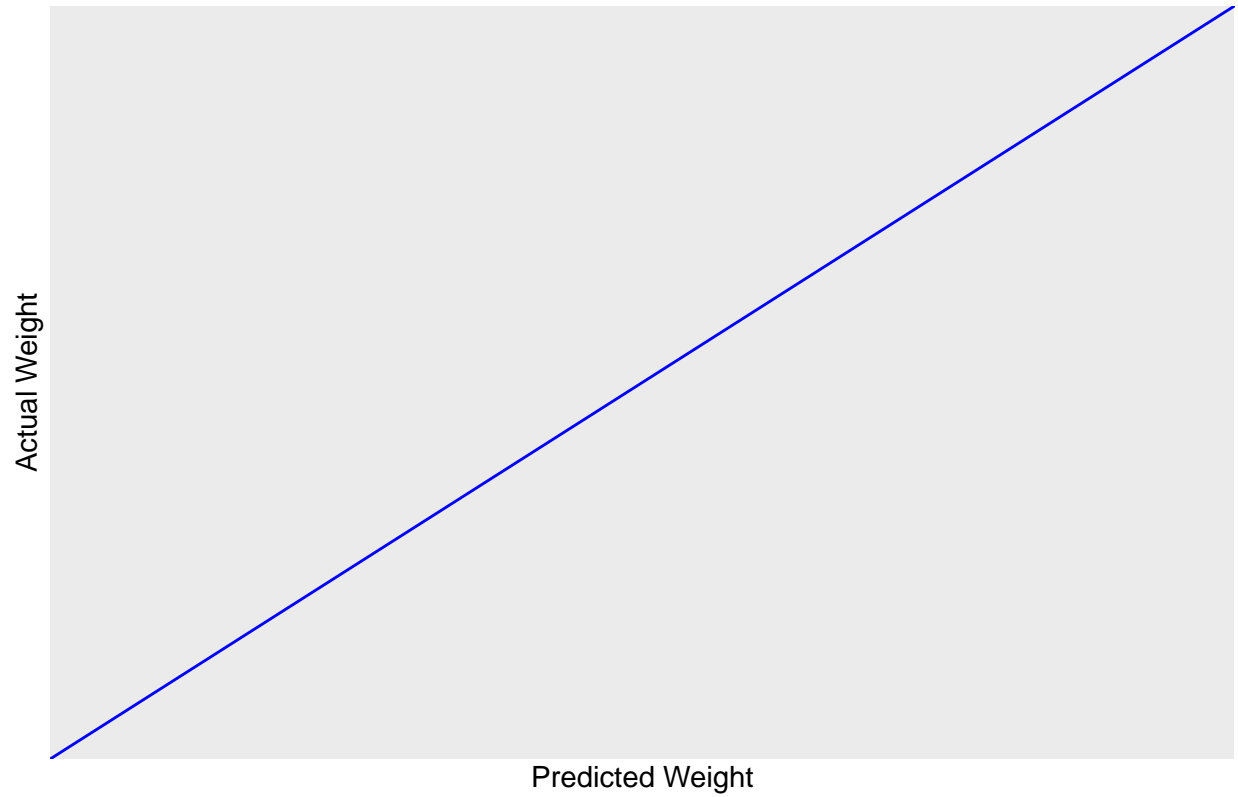
```
## Mean Squared Error: NaN \n
```

```
cat("R-squared:", r_squared, "\\n")
```

```
## R-squared: NaN \n
```

Step 6: Visualize Results

Predicted vs Actual

Predicted vs Actual Weight

Residual Plot