

My Tutorial

Introduction

In this tutorial, we refactored the code into separate scripts corresponding to each section. The dataset comes from the `palmerpenguins` package, which contains measurements of penguins from three species. All results shown below are generated from scripts and saved outputs.

Load Libraries and Data

We used the `palmerpenguins`, `tidyverse`, and `tidymodels` packages.
Initial data cleaning (e.g., removing NA values) was handled in the scripts.

To inspect the data structure:

Rows..333

Columns: 8

\$ species Adelie, Adelie, Adelie, Adelie, Adelie, A...

\$ island Torgersen, Torgersen, Torgersen, Torgersen, ... \$ bill_length_mm 39.1, 39.5, 40.3, 36.7, 39.3, 38.9, 39.2, 41.1, 38.6... \$ bill_depth_mm 18.7, 17.4, 18.0, 19.3, 20.6, 17.8, 19.6, 17.6, 21.2... \$ flipper_length_mm 181, 186, 195, 193, 190, 181, 195, 182, 191, 198, 18... \$ body_mass_g 3750, 3800, 3250, 3450, 3650, 3625, 4675, 3200, 3800... \$ sex male, female, female, female, male, femal... | \$ year 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007...
|

Methods

In this section, we perform exploratory data analysis (EDA) and prepare the data for modeling.

mean_bill_length	mean_bill_depth
43.99279	17.16486

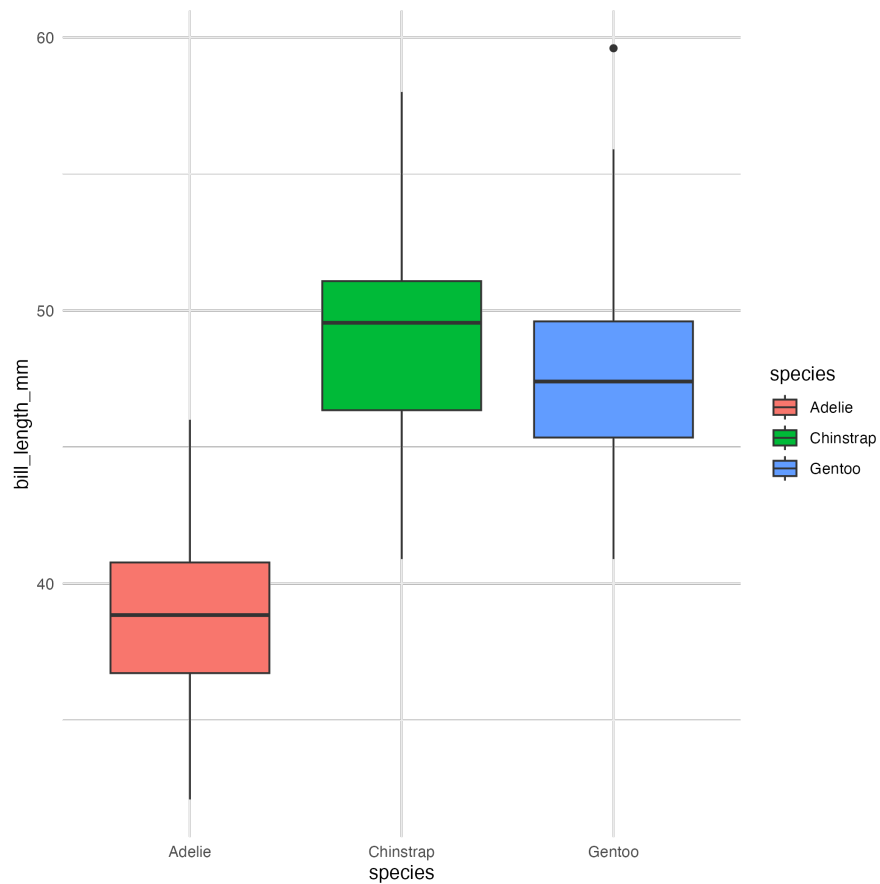


Figure 1: Bill Length Boxplot

Model

We will fit a classification model using `tidymodels` to predict the species of a penguin based on its physical characteristics.

Results

We evaluate the performance of the model using the test dataset.

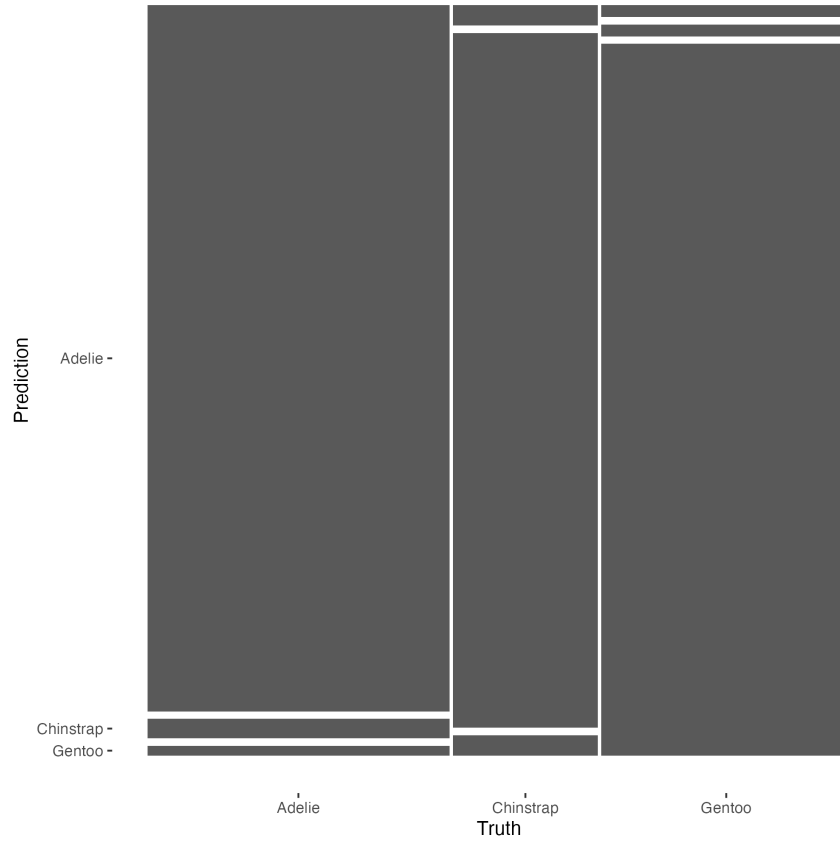


Figure 2: Confusion Matrix for our model

Conclusion

In this tutorial, we:

- Loaded and cleaned the `palmerpenguins` dataset.
- Performed exploratory data analysis.
- Built a k-Nearest Neighbors classification model using `tidymodels`.
- Evaluated the model's performance.