

Data Partitioning and Modeling (Catch the Pink Flamingo)

Analysis of combined_data.csv | Nyan Lynn Htet | 8 Sep 2024

The data was partitioned into train and test datasets.

The **training** data set was used to create the decision tree model.

The trained model was then applied to the **test data** dataset.

This is important because partitioning the data set into **training** and **test data** allows me to verify the output accuracy of the trained model.

When partitioning the data using sampling, it is important to set the random seed because it ensures that you get the same results every time you perform the partitioning.

A screenshot of the resulting decision tree can be seen below:

