

Data Partitioning and Modeling

The data was partitioned into train and test datasets.

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

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

This is important because **preventing model overfitting**.

When partitioning the data using sampling, it is important to set the random seed because **the train/test dataset will not change regardless rerun all workflow**.

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

