**Data Partitioning and Modeling**

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 dataset.

This is important because…

Partitioning the data set into training and test data allows us to verify the accuracy of the trained model. If we used all the data for training the model, we would have no idea how the model performs.

When partitioning the data using sampling, it is important to set the random seed because…

A random seed will partition the data set consistently. This allows you to obtain reproducible results each time you run the partition. When validating the accuracy of different models, you need to keep the data used to test the models consistent. If the test data is inconsistent, you can’t compare the accuracy between models.

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

