

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... in order to test the generalisability of the decision tree model, it needs to be tested on unseen data, i.e data it was NOT trained on.

When partitioning the data using sampling, it is important to set the random seed because... the partitions made have to be reproducible for experimental purposes, such as comparisons with other methods or other people's approaches.

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

