**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… **we must train the model with some data , and test the model with other data. We must not train the model and test it with the same data to avoid overfitting.**

When partitioning the data using sampling, it is important to set the random seed because… **we want repeatable result, meaning to get same result when repeating prediction with the same data. So we can compare our results and evaluate this exercise.**

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



