

2.2 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 train data set is used in creating the decision tree model, the apply the model to the test data set, which is not used to train the mode then we can see the accuracy of the model.

When partitioning the data using sampling, it is important to set the random seed because it can get the same data partitions every time the node is executed.

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



