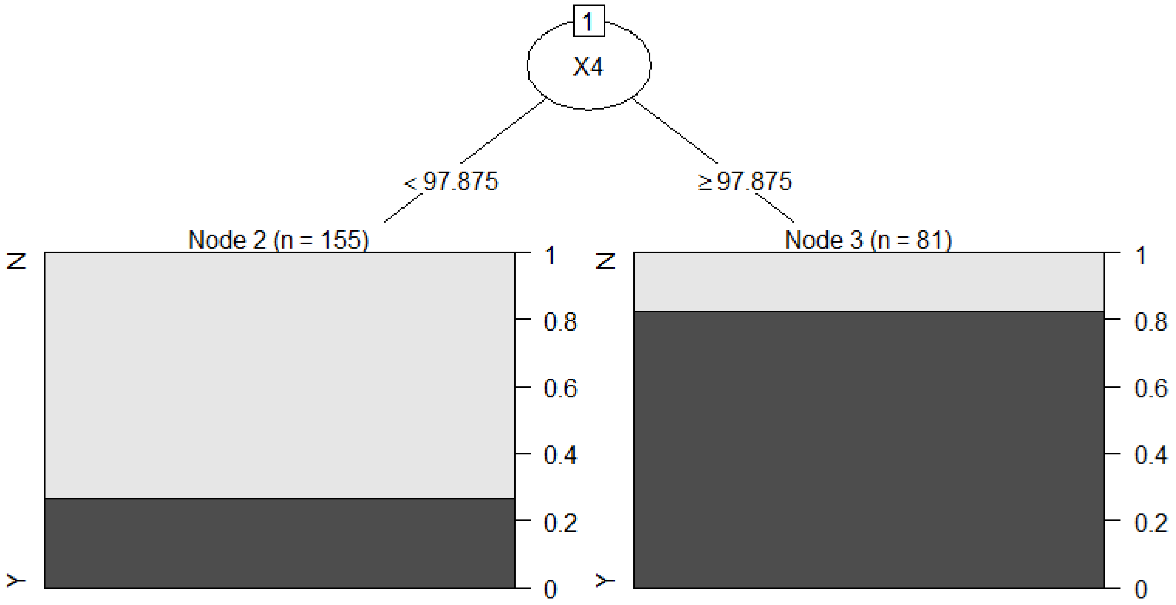
9 models with Data Imputation:

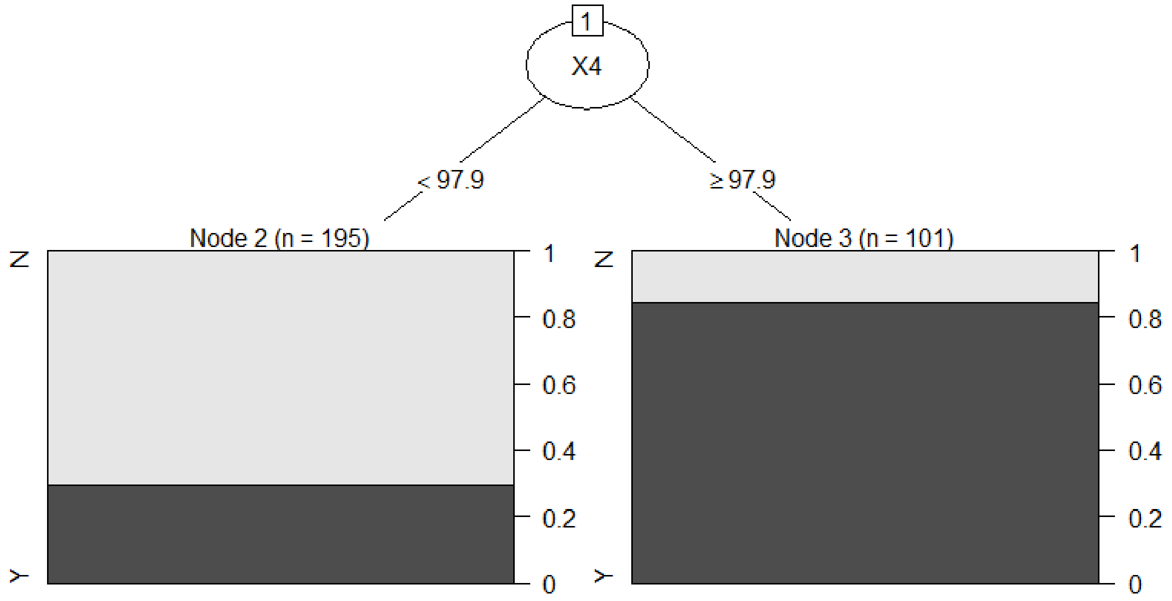
X vs Response results

Dataset is imputed using mice library and following are the results

Model pruned for training data:



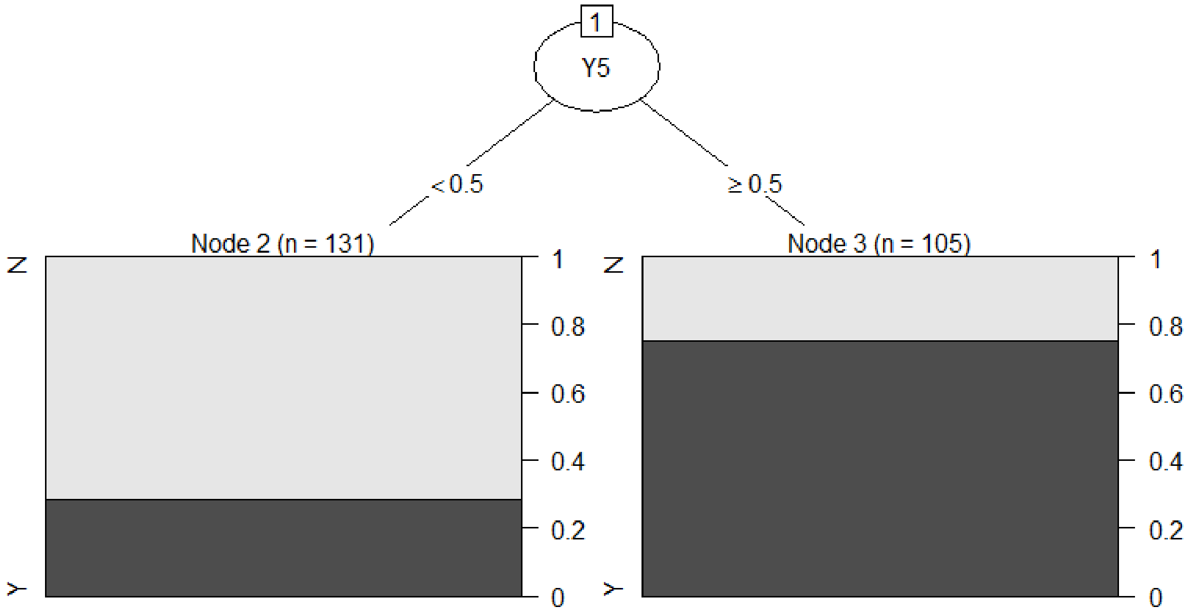
Model pruned with whole dataset:



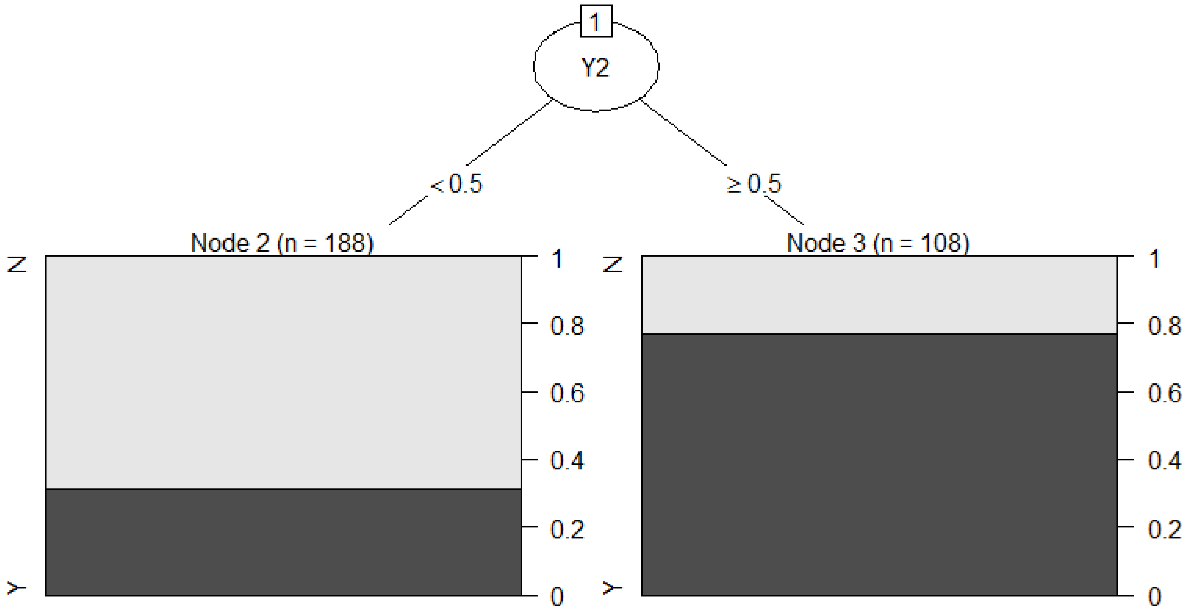
Y vs Response results

Dataset is imputed using mice library and following are the results

Model pruned with training data:



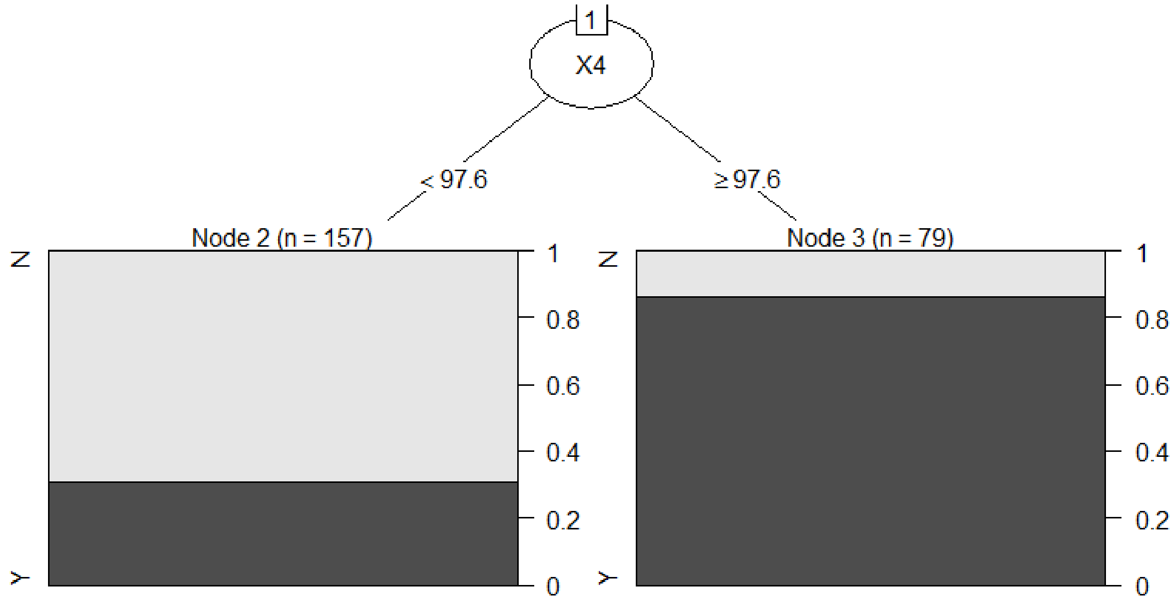
Model pruned with whole dataset:



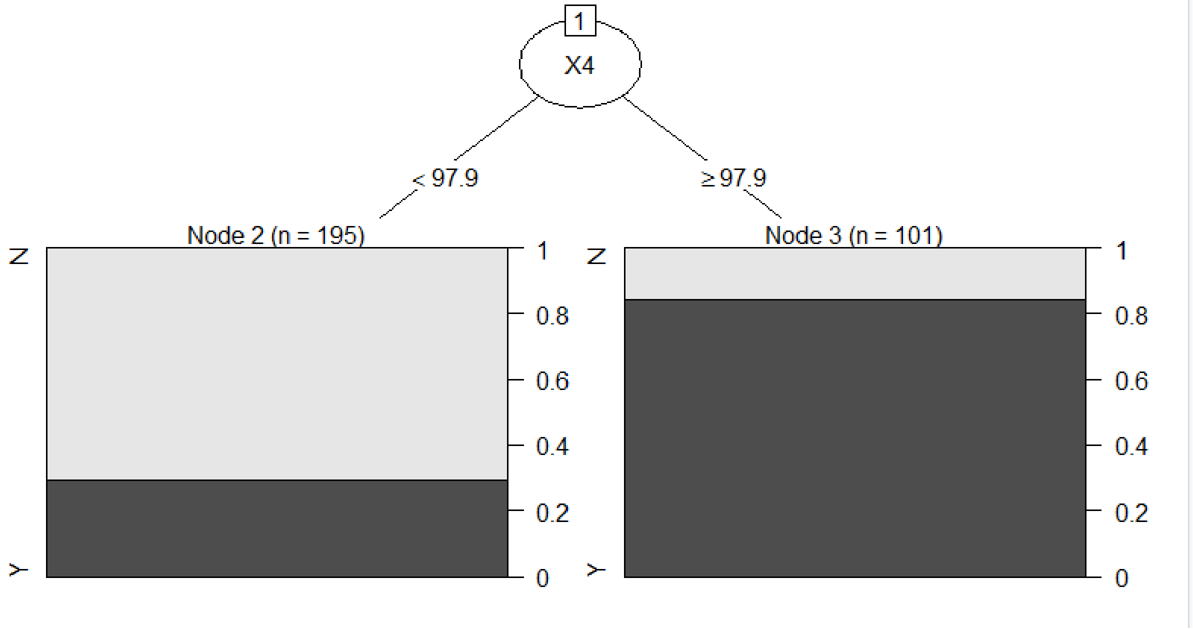
XY vs Response results:

Dataset is imputed using mice library and following are the results

Model pruned with training dataset:



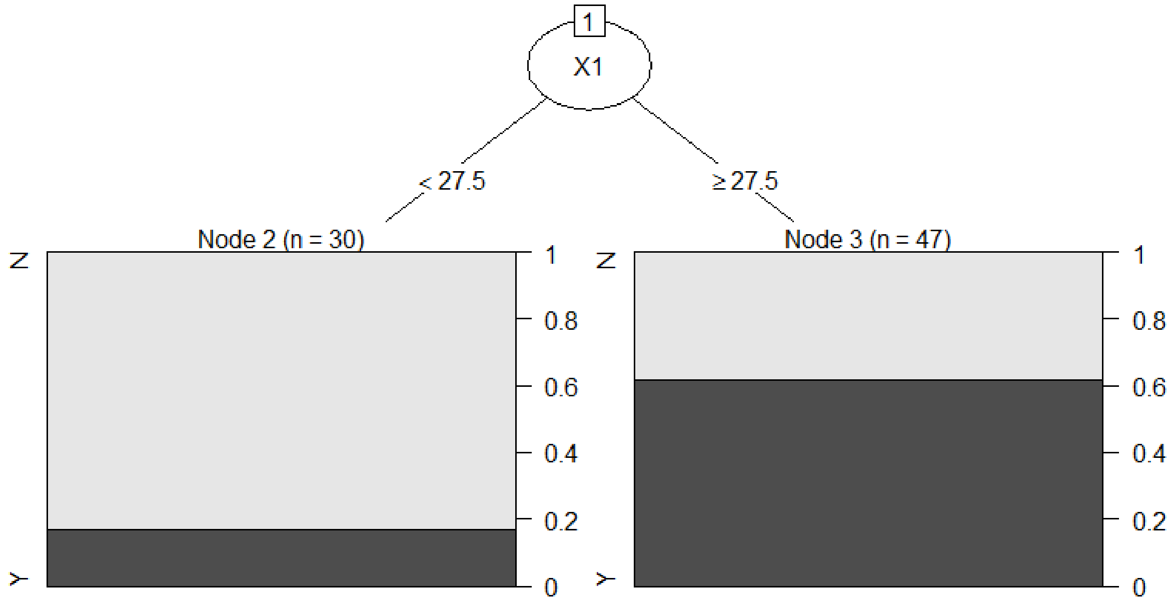
Model pruned with whole dataset:



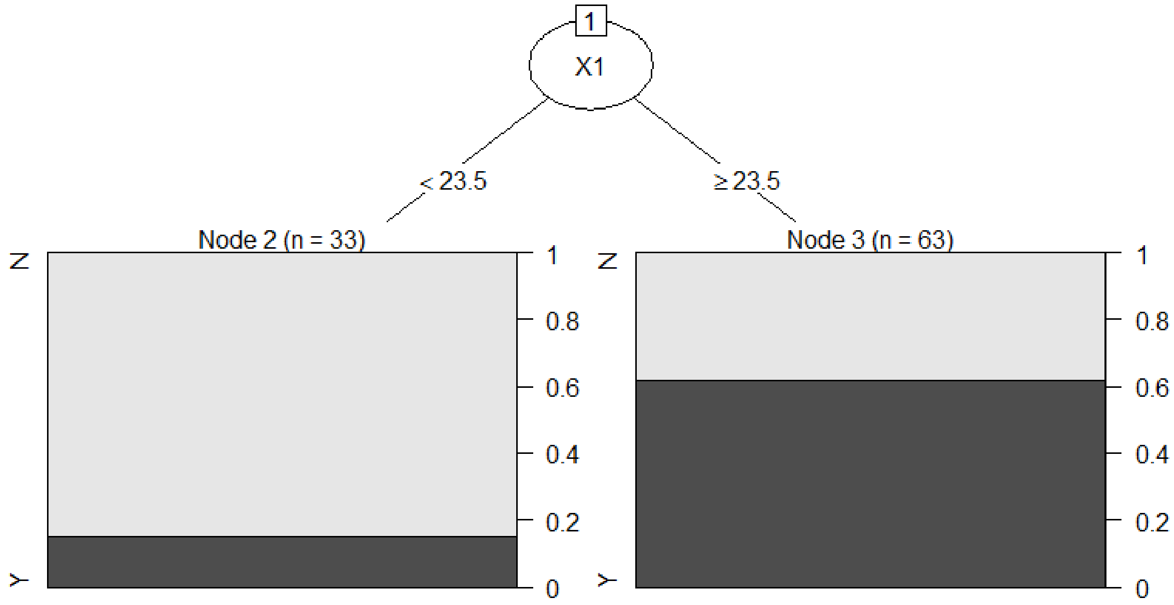
X-0 vs Response results

Dataset is imputed using mice library and following are the results

Model pruned with training dataset:



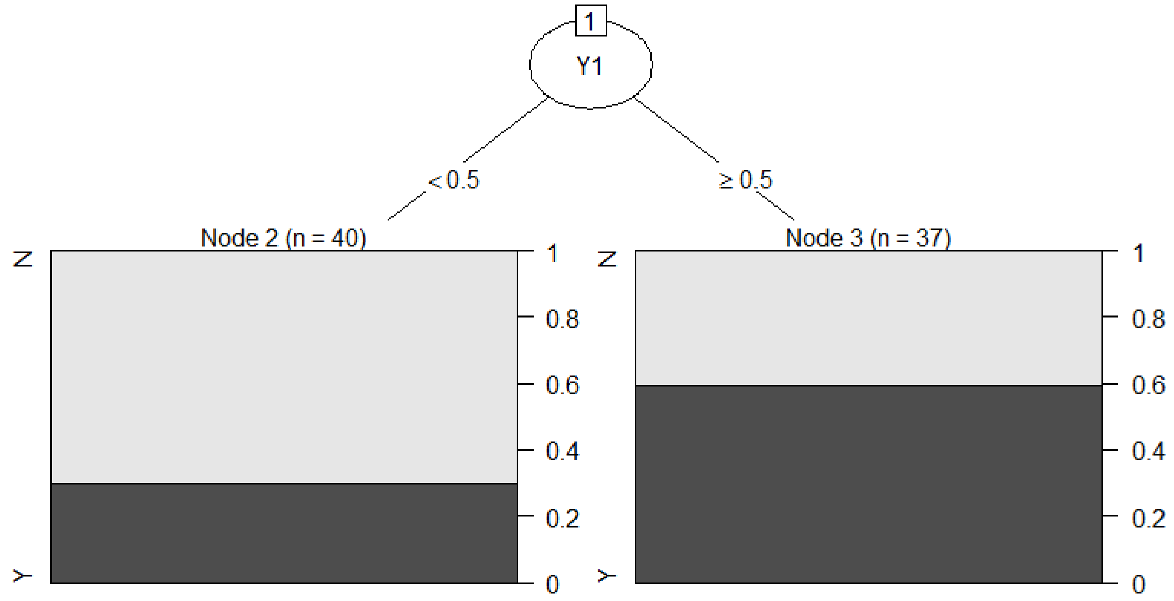
Model pruned with whole dataset:



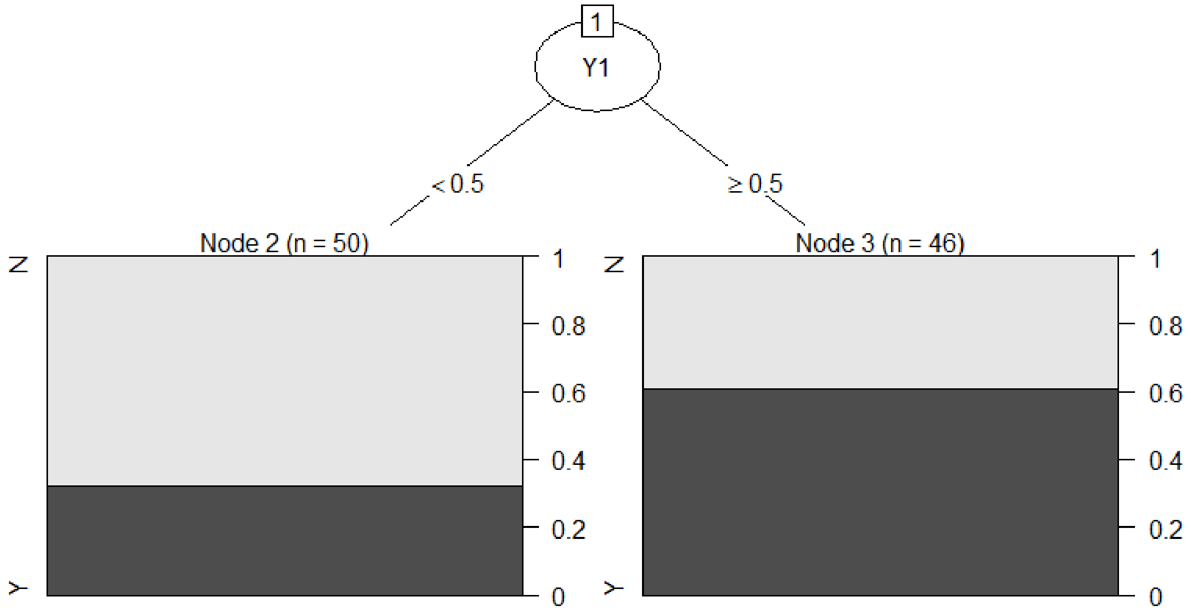
Y-0 vs Response results

Dataset is imputed using mice library and following are the results

Model pruned with training dataset:



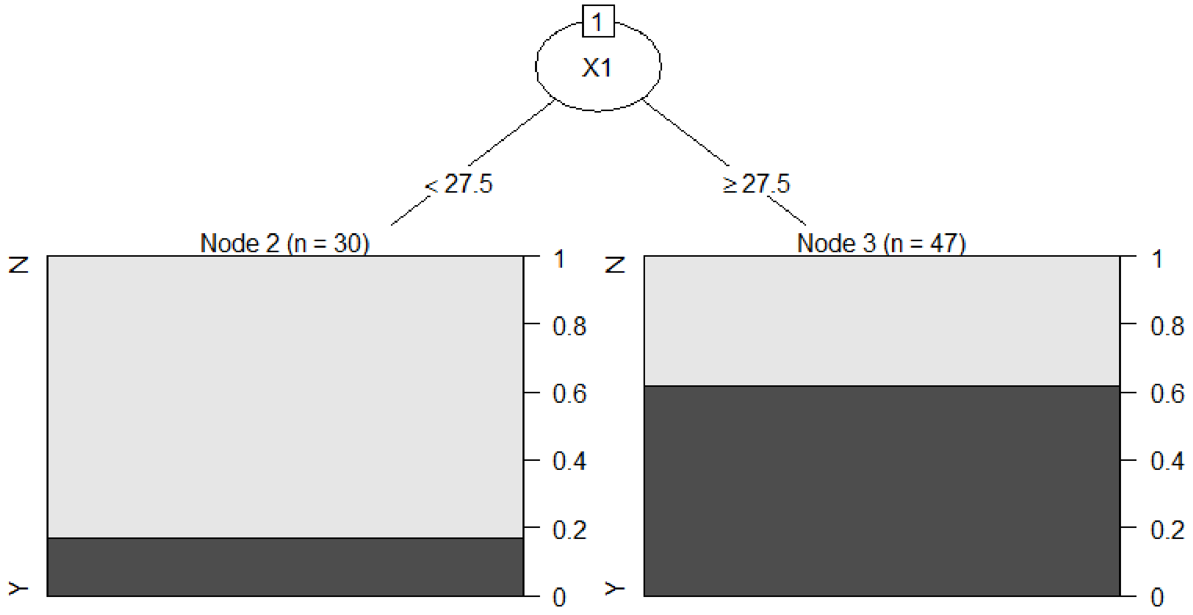
Model pruned on whole dataset:



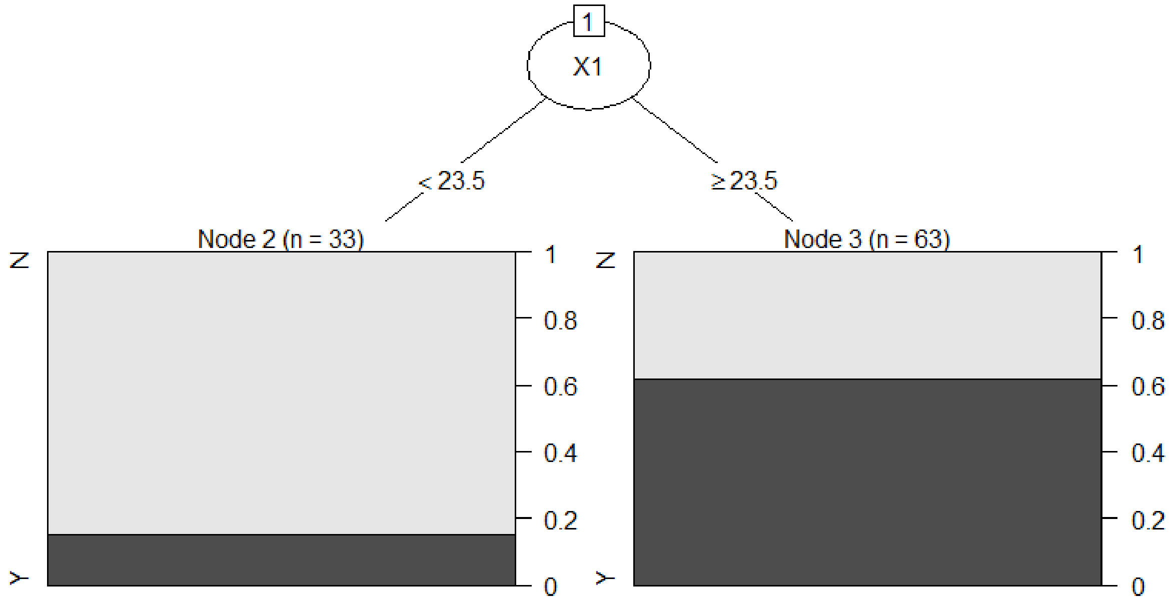
XY-0 vs Response results

Dataset is imputed using mice library and following are the results

Model pruned with training data:



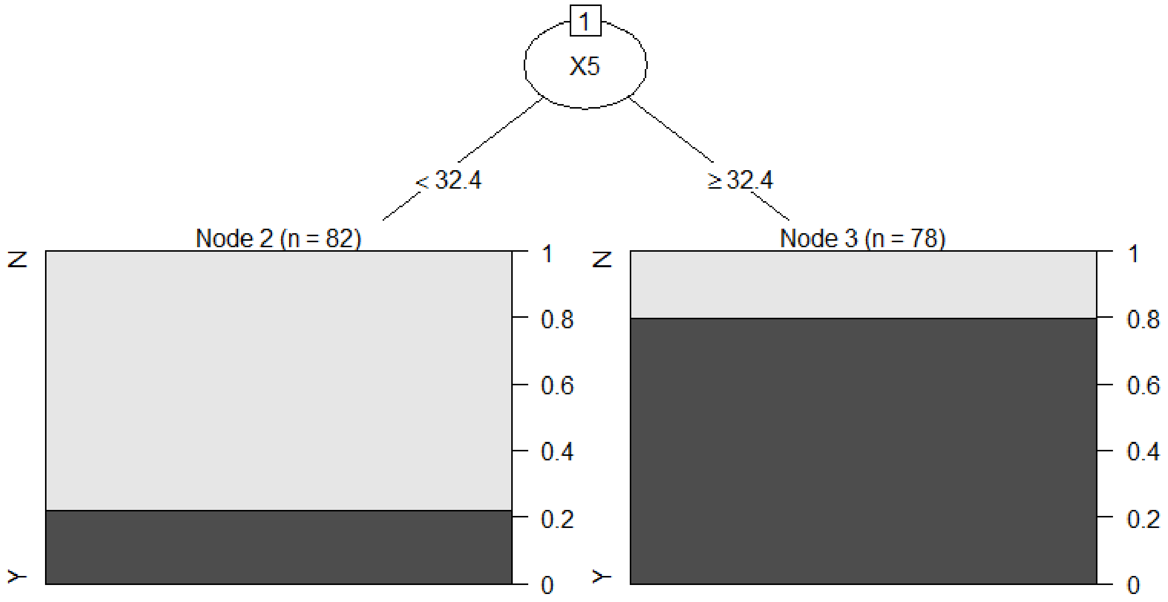
Model pruned with whole dataset:



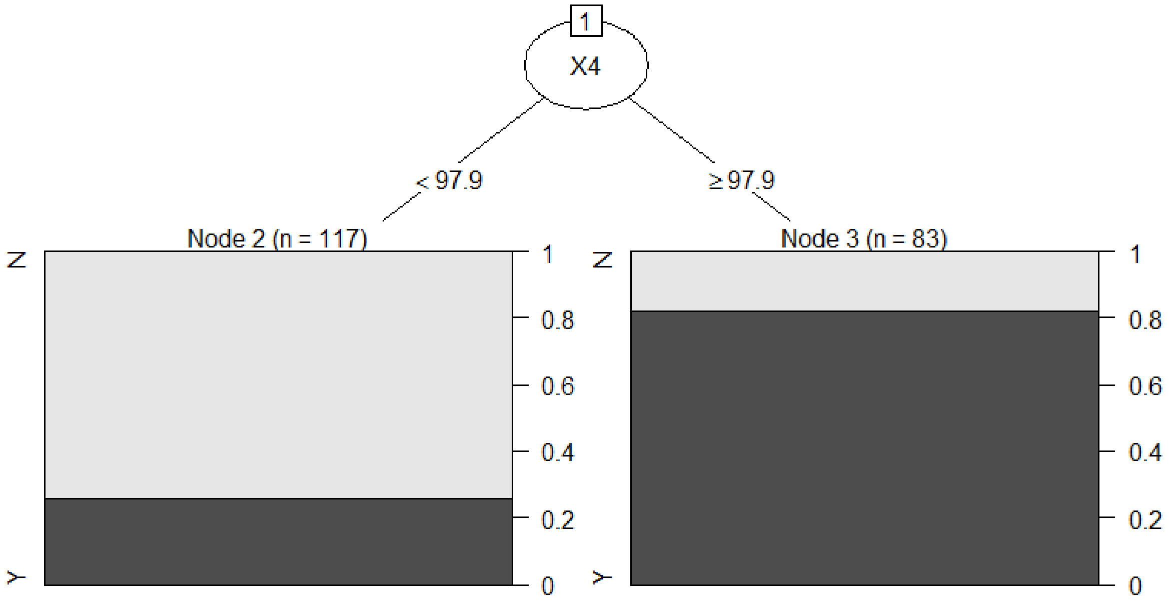
X-1 vs Response results

Dataset is imputed using mice library and following are the results

Model pruned with training dataset :



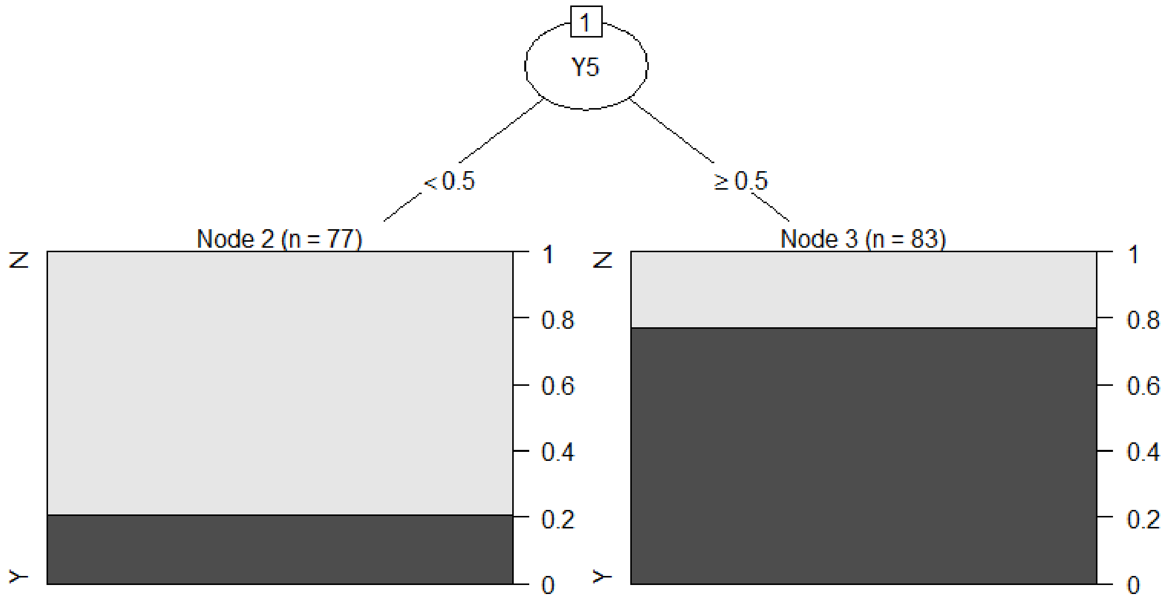
Model pruned on whole dataset :



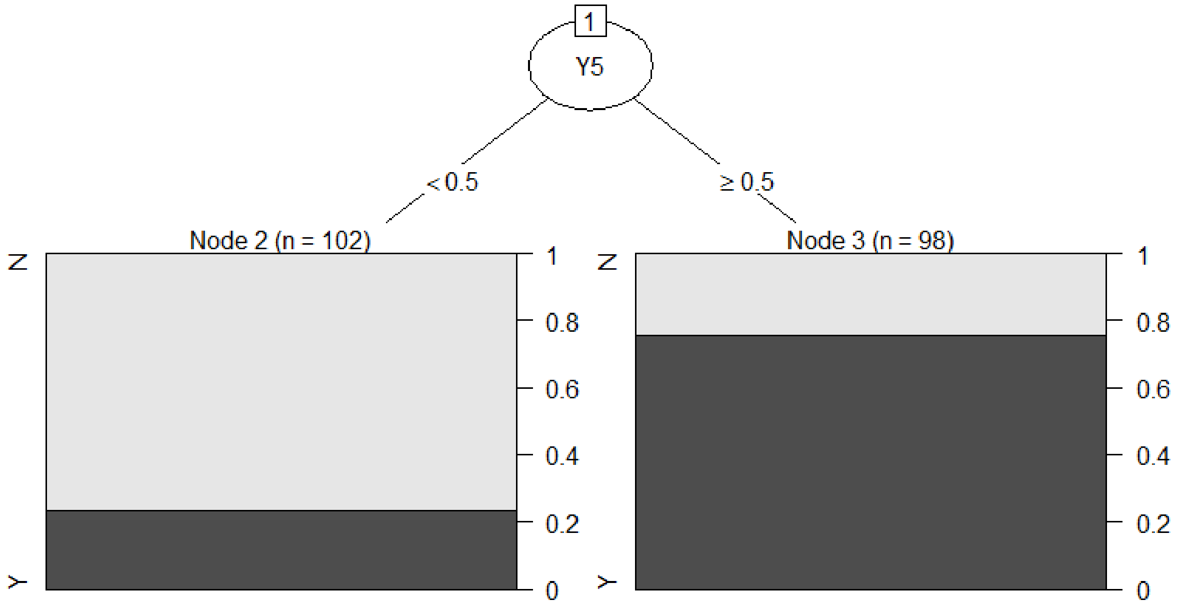
Y-1 vs Response results

Dataset is imputed using mice library and following are the results

Model pruned on training dataset:



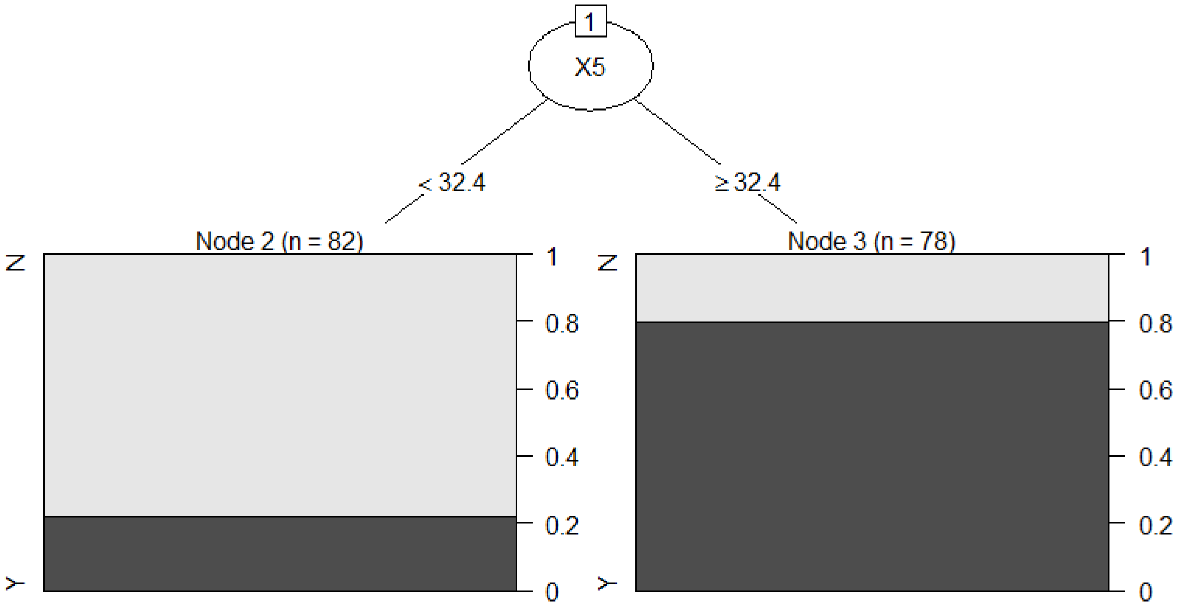
Model pruned on whole dataset:



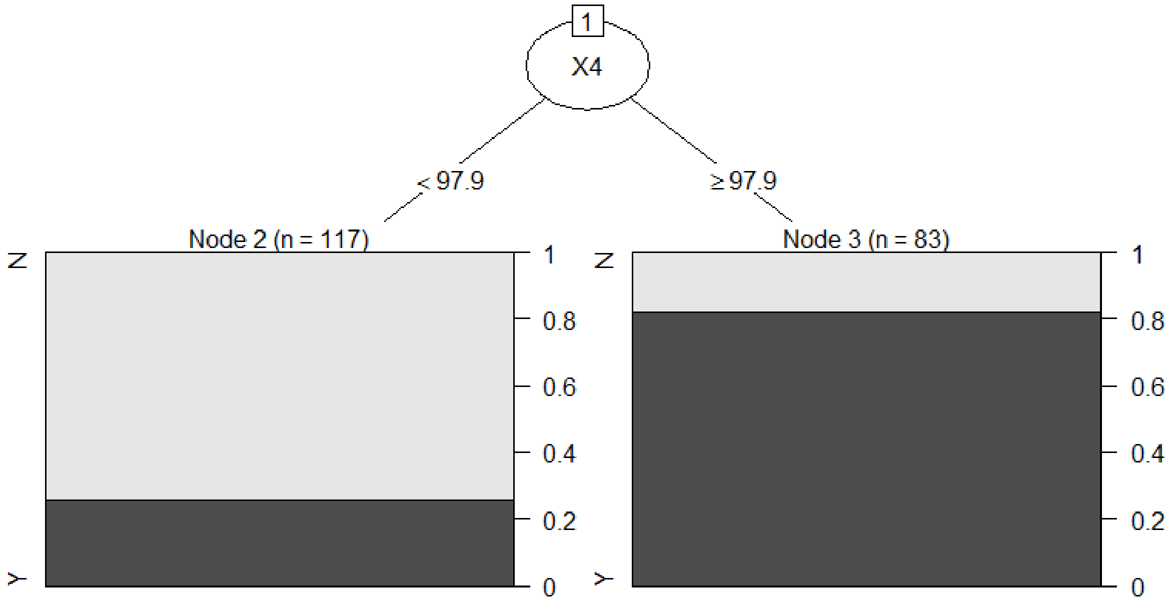
XY-1 vs Response results

Dataset is imputed using mice library and following are the results

Model pruned on training dataset:



Model pruned on whole dataset :



Conclusion:

Here the dataset columns of Xs and Ys are imputed with mean using the “mice” library. 2 pruned models are prepared : One is by splitting the dataset into training and test data. The other is by taking he whole dataset.

The Response is in the form of 1 and 0 in the original dataset. This is converted to Y and N and named as **Target column**. Then dataset is **split into 80% and 20% for training and validation**. Pruned model is developed and prediction is run on test dataset and the results are compared with original expected responses in the test data. The comparison is represented in the form of **confusion matrix** .

The best model is then predicted on the basis of %Accuracy which is calculated by {true positives+ true negatives }/{[true positives+true negatives+false positives+false negatives]}

I have found the **model X-1 vs Response working best with average %Accuracy around 75%.**

The condition for best model **(rel error + xstd < xerror)** fits the best in this model as the difference between the LHS and RHS of the above expression is the maximum in this model.

```
> print(DT_Model3$cptable)
         CP nsplit rel error    xerror       xstd
1 0.5408163      0 1.0000000 1.0000000 0.07213932
2 0.0100000      1 0.4591837 0.5612245 0.06443536
```