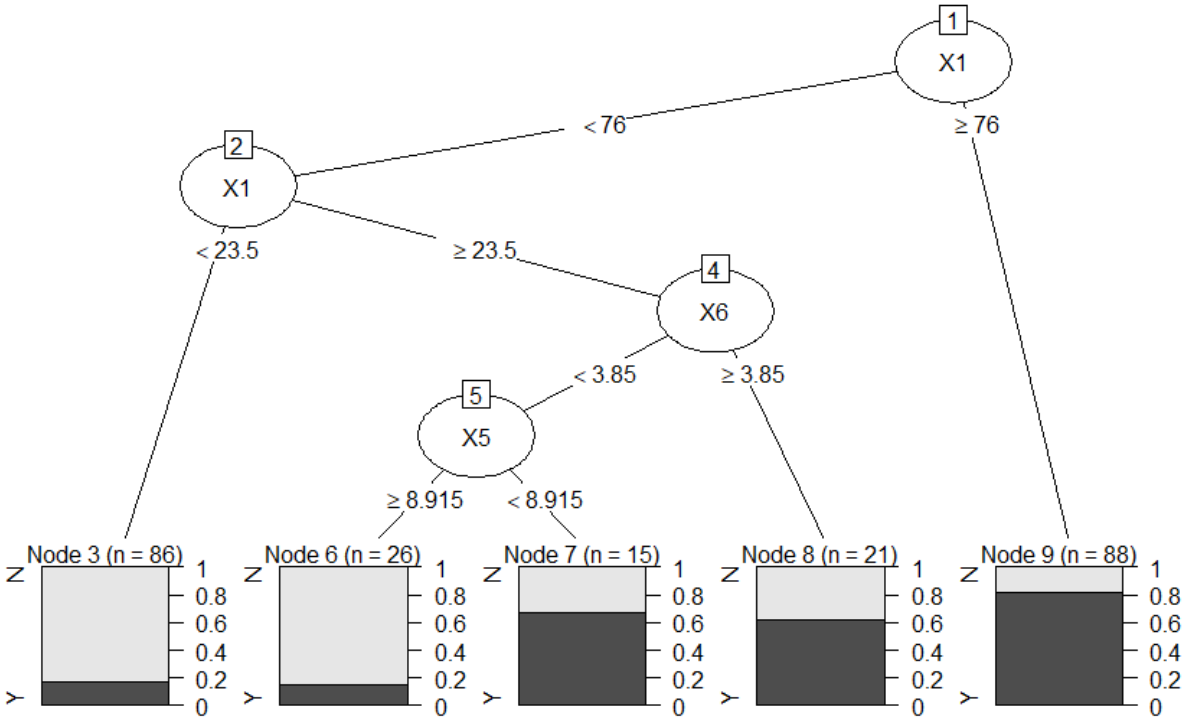
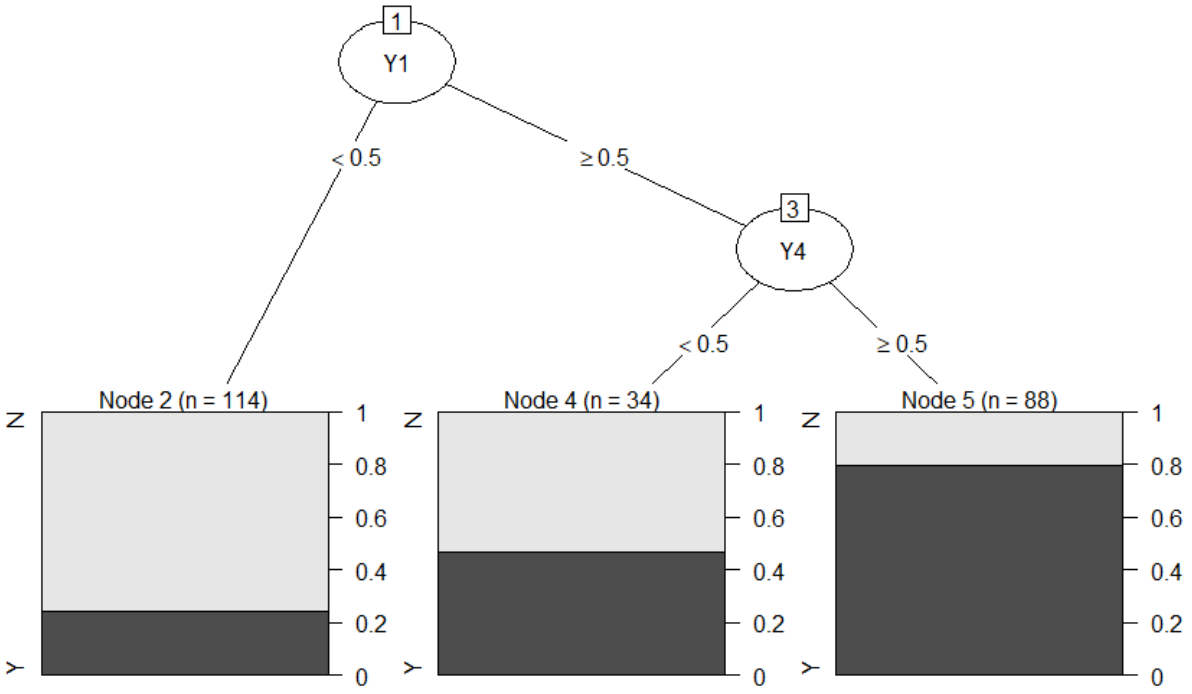
Project Report for CA 1

By Arambakam Mukesh - 19301497



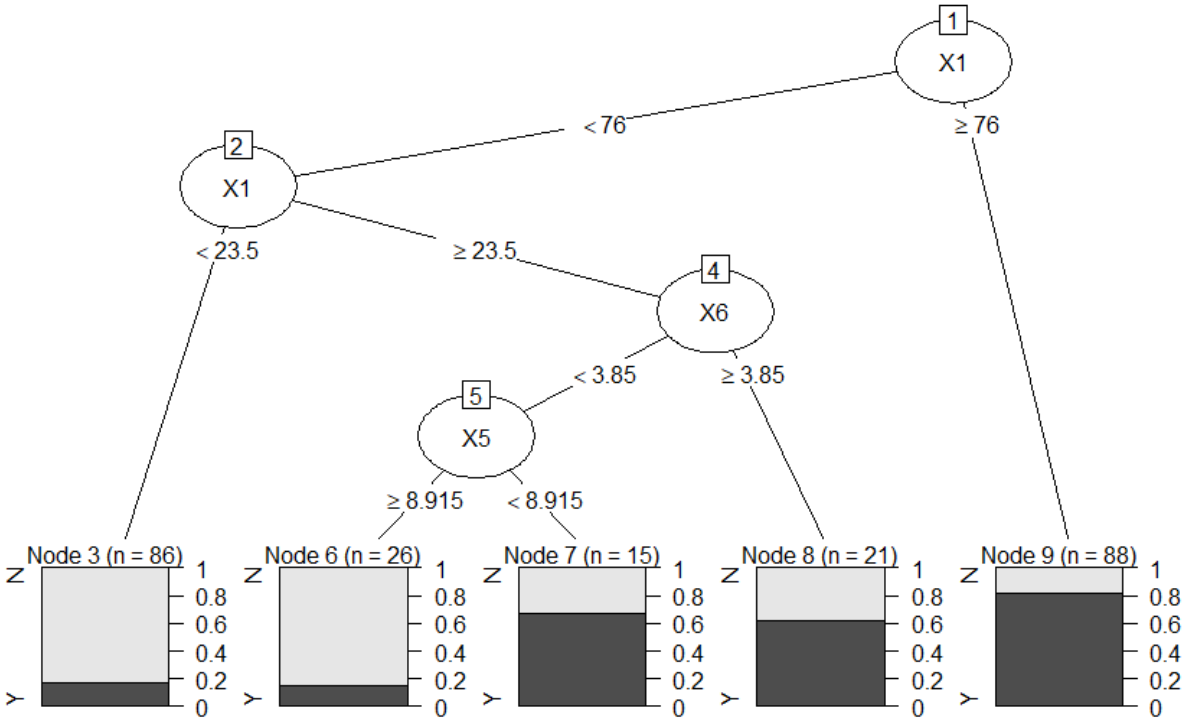
Plot 1 - DT over only X

Plot 1 represents the Decision Tree over the entire data set but with the Predictors though X1-X7. This DT predicts the data with an accuracy of 65% (0.65).



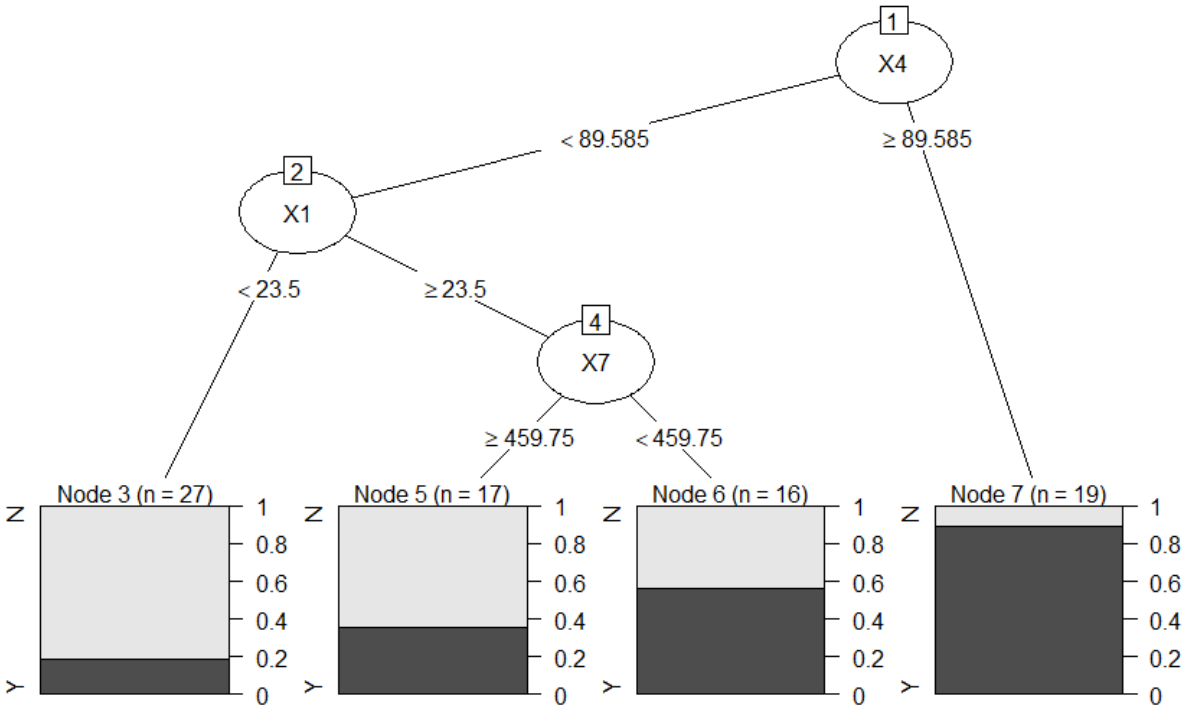
Plot 2 - DT over only Y

Plot 2 represents the Decision Tree over the entire data set but with the Predictors though Y1-Y7. This DT predicts the data with an accuracy of 72% (0.72).



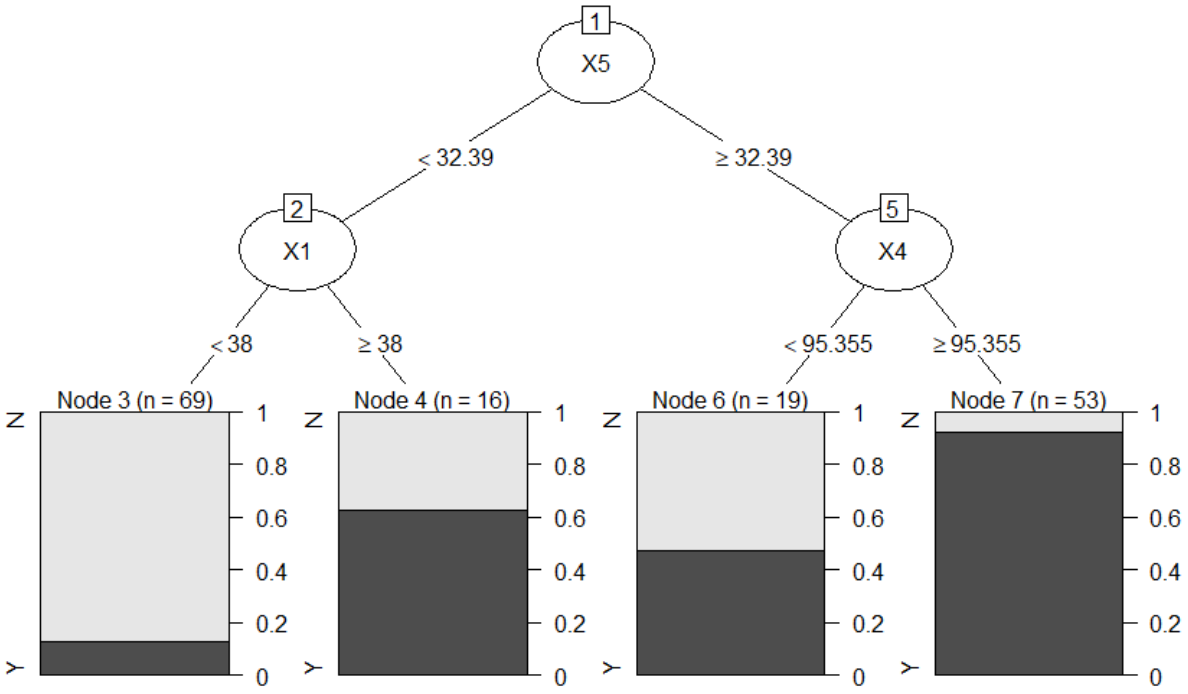
Plot 3 - DT over X and Y

Plot 3 represents the Decision Tree over the entire data set but with the Predictors though X1-X7 and Y1-Y7. This DT predicts the data with an accuracy of 65% (0.65).



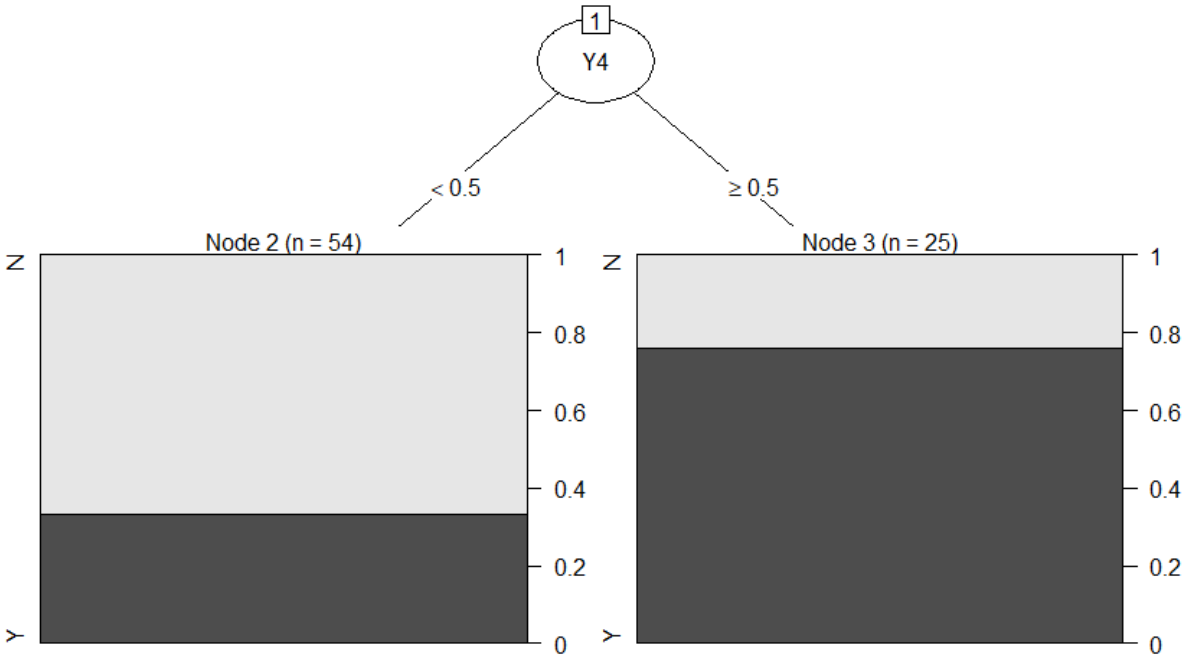
Plot 4 - DT over X with Group 0

Plot 4 represents the Decision Tree over Group 0 set but with the Predictors though X1-X7. This DT predicts the data with an accuracy of 73% (0.73).



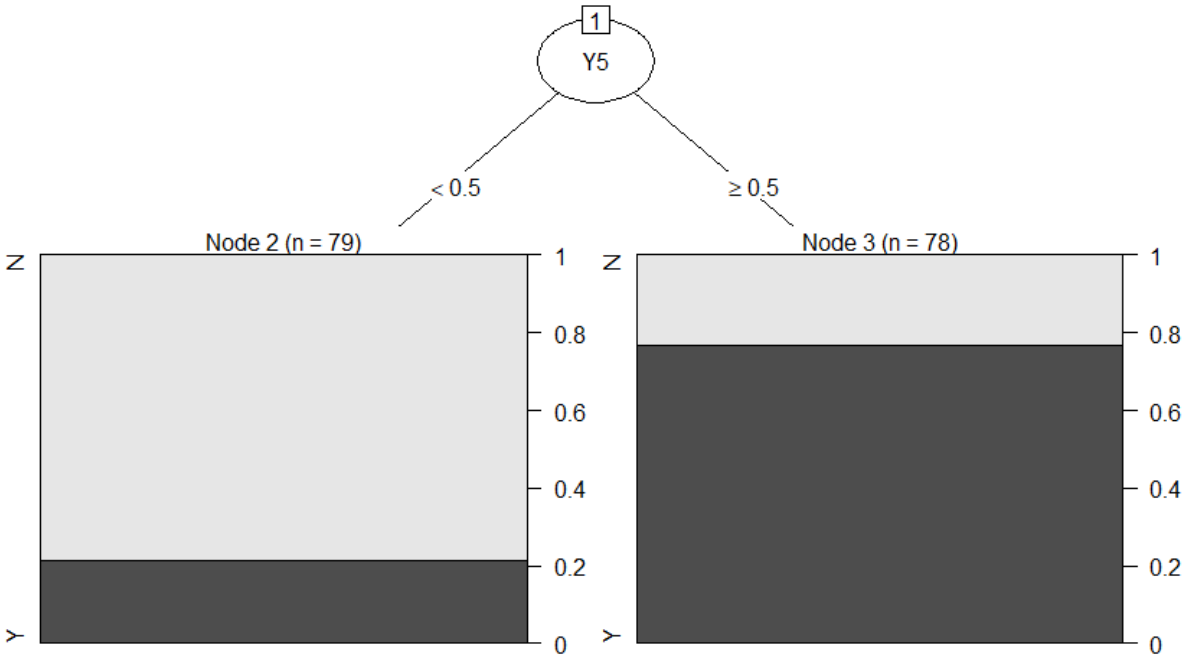
Plot 5 - DT over X with Group 1

Plot 5 represents the Decision Tree over Group 1 set but with the Predictors though X1-X7. This DT predicts the data with an accuracy of 70% (0.7).



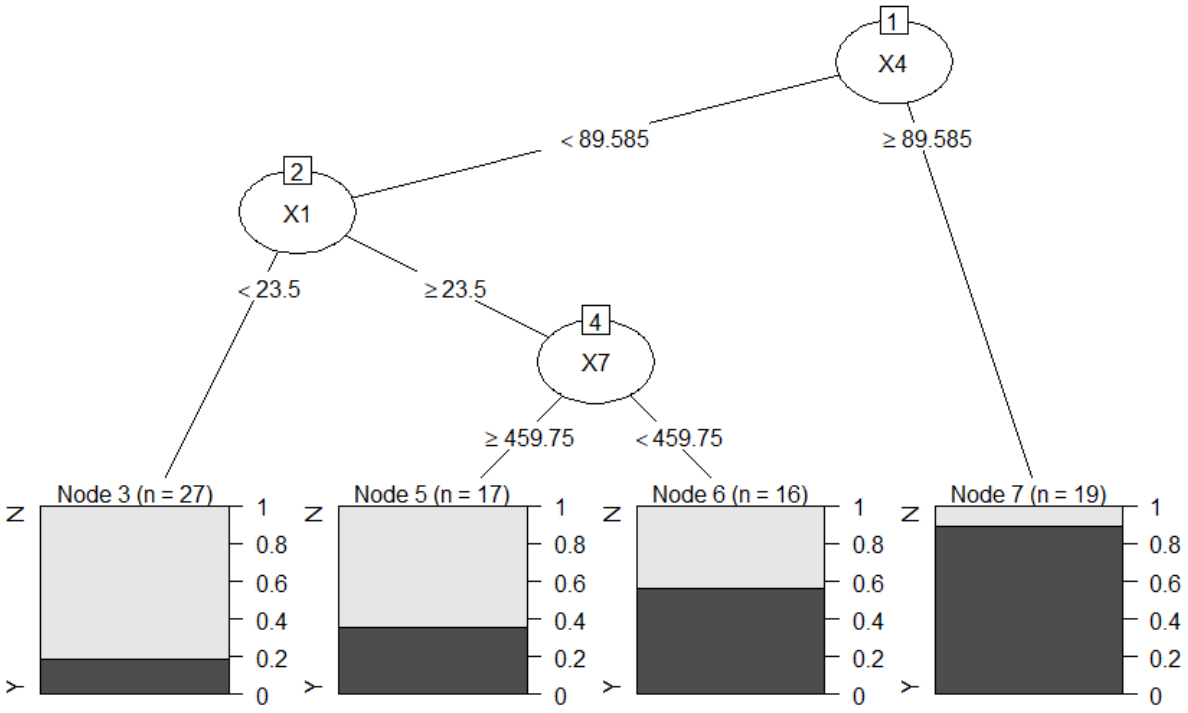
Plot 6 - DT over Y with Group 0

Plot 6 represents the Decision Tree over Group 0 set but with the Predictors though Y1-Y7. This DT predicts the data with an accuracy of 70% (0.7).



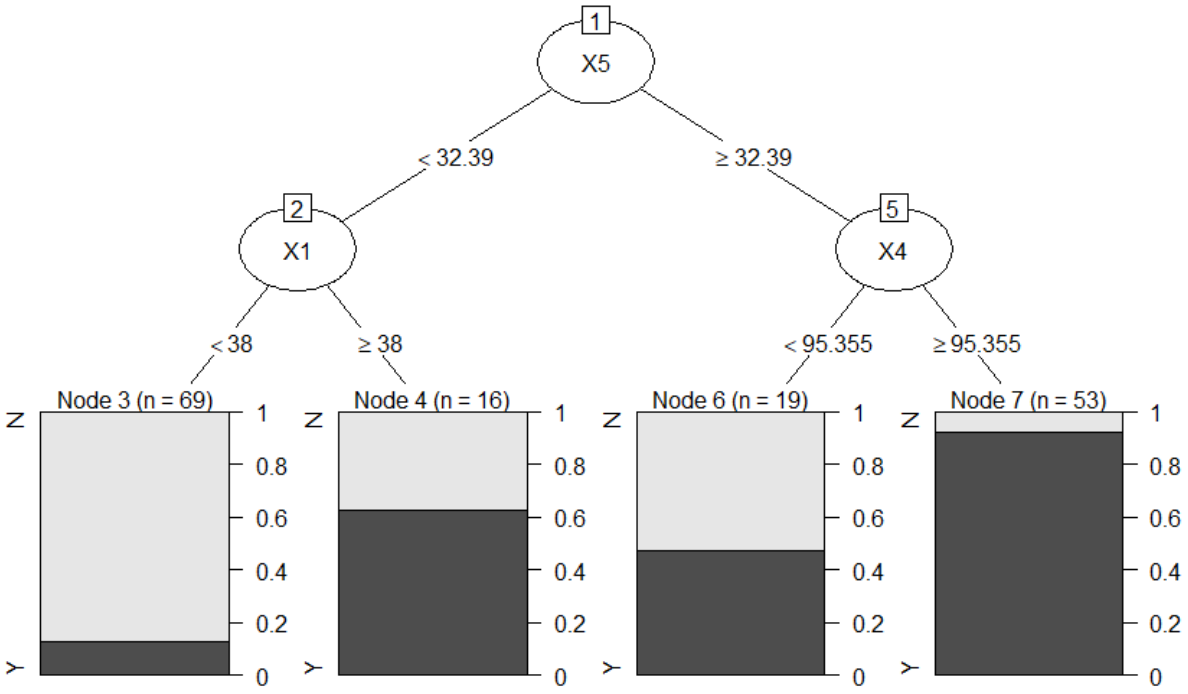
Plot 7 - DT over Y with Group 1

Plot 7 represents the Decision Tree over Group 1 set but with the Predictors though Y1-Y7. This DT predicts the data with an accuracy of 70% (0.7).



Plot 8 - DT over X and Y with Group 0

Plot 8 represents the Decision Tree over Group 0 set but with the Predictors though X1-X7 and Y1-Y7. This DT predicts the data with an accuracy of 73% (0.73).



Plot 9 - DT over X and Y with Group 1

Plot 9 represents the Decision Tree over Group 1 set but with the Predictors though X1-X7 and Y1-Y7. This DT predicts the data with an accuracy of 70% (0.7).

Conclusion:

The best Decision Tree generated is the DT generated over **Group 0 with the Predictors X1-X7** as it has the highest **accuracy of 73%** - see **Plot 4** for the DT. The below is the Decision Tree's summary, indicating the splits:

```
n= 79

node), split, n, loss, yval, (yprob)
      * denotes terminal node

 1) root 79 37 N (0.5316456 0.4683544)
   2) X4< 89.585 60 20 N (0.6666667 0.3333333)
     4) X1< 23.5 27  5 N (0.8148148 0.1851852) *
     5) X1>=23.5 33 15 N (0.5454545 0.4545455)
      10) X7>=459.75 17  6 N (0.6470588 0.3529412) *
      11) X7< 459.75 16  7 Y (0.4375000 0.5625000) *
   3) X4>=89.585 19  2 Y (0.1052632 0.8947368) *
```

Though the DT generated over Group 0 with Predictors X1-X7 and Y1-Y7 also has an accuracy of 73% - see Plot 8 for the DT – it requires an additional predictor to result into the same DT as shown in Plot 4. Both Plot 4 and 8 have the same DT thought different predictors were used. The final DT in the Plot uses the same.

The code for this can be found on my GitHub, please find the link to the repo below.

https://github.com/mukeshmk/r-project