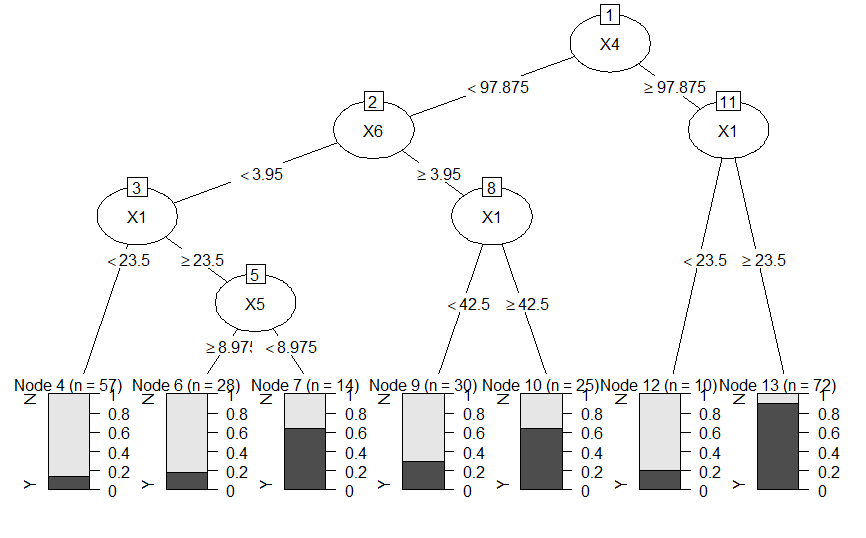
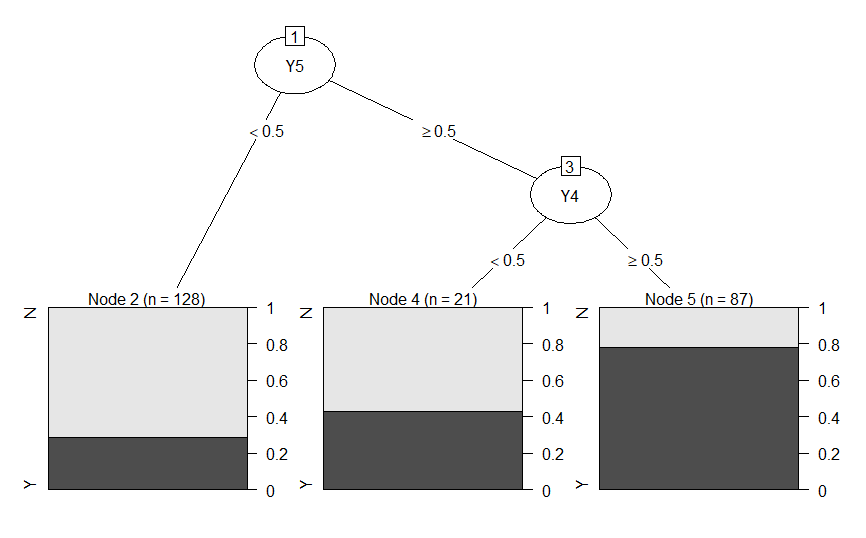
**Project Report for CA 1**

By Arambakam Mukesh - 19301497



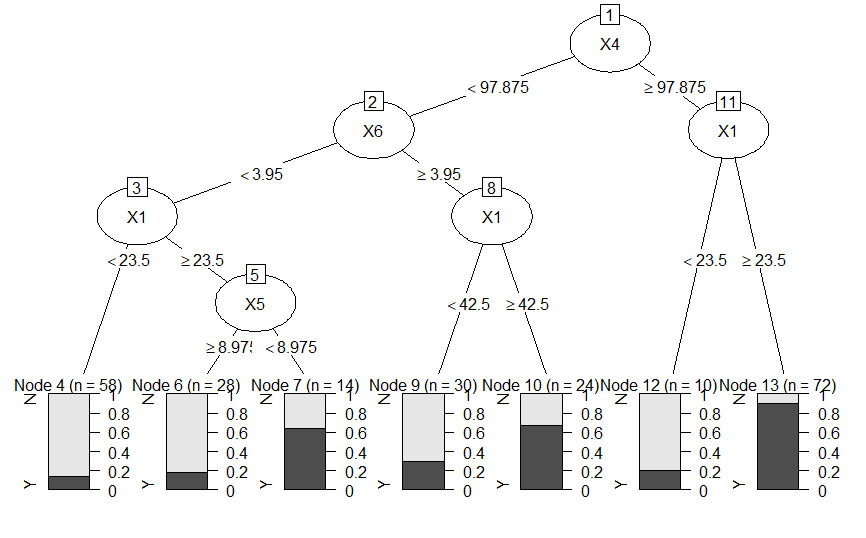
Plot 1 - DT over only X

Plot 1 represents the Decision Tree over the entire data set but with the Predictors though X1-X7. This DT predicts with an accuracy of 80% (0.8).



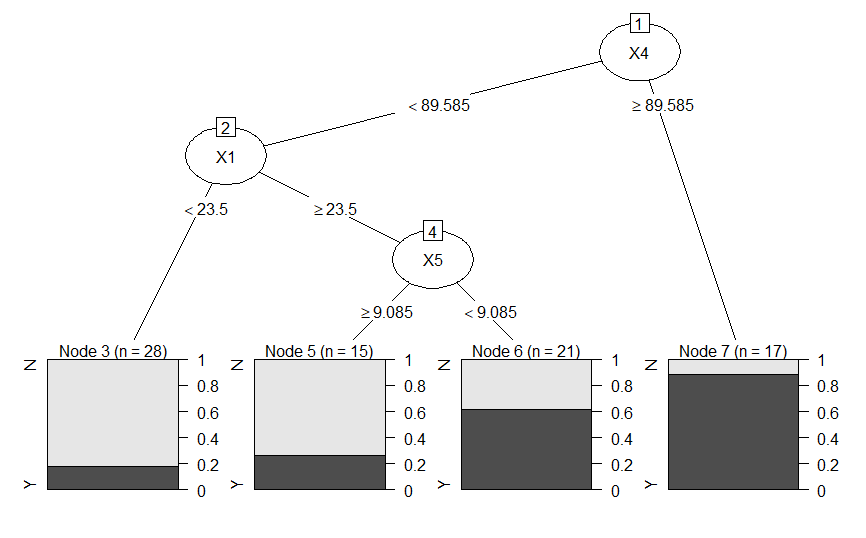
Plot 2 - DT over only Y

Plot 2 represents the Decision Tree over the entire data set but with the Predictors though Y1-Y7. This DT predicts with an accuracy of 80% (0.8).



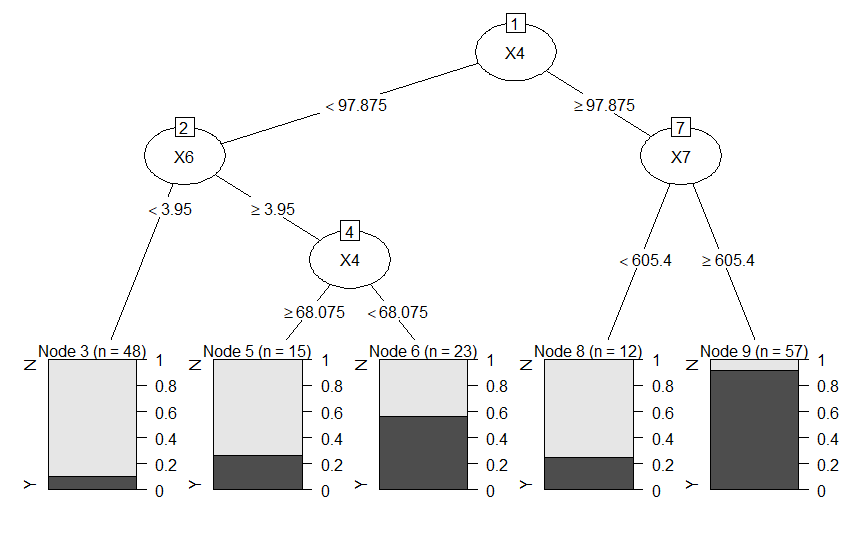
Plot 3 - DT over X and Y

Plot 3 represents the Decision Tree over the entire data set but with the Predictors though X1-X7 and Y1-Y7. This DT predicts with an accuracy of 80% (0.8).



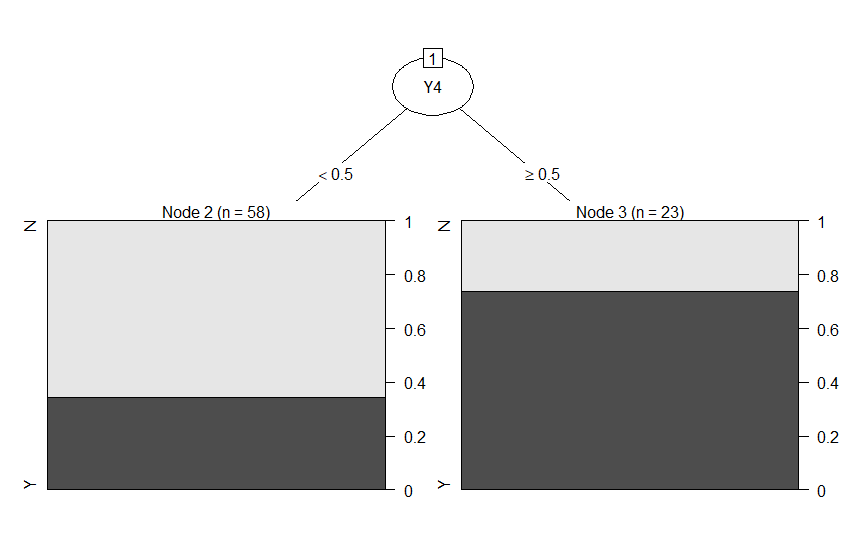
Plot 4 - DT over X with Group 0

Plot 4 represents the Decision Tree over Group 0 set but with the Predictors though X1-X7. This DT predicts with an accuracy of 77% (0.77).



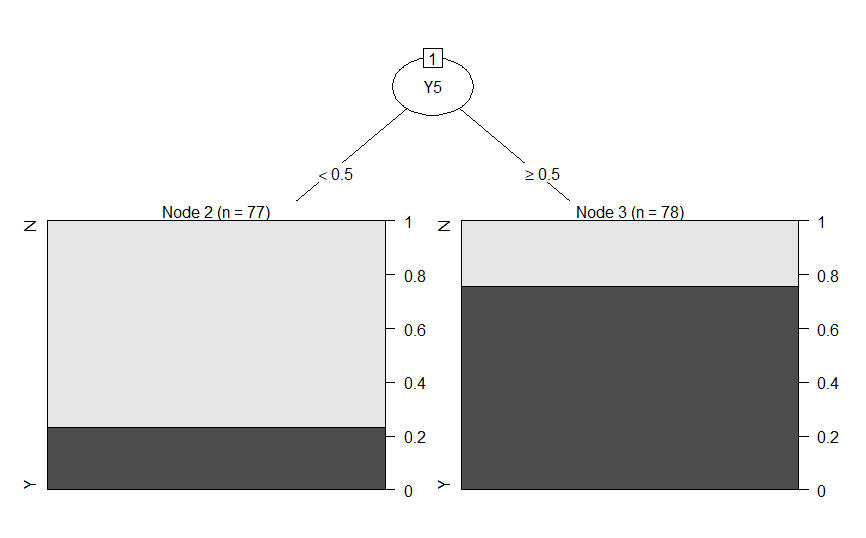
Plot 5 - DT over X with Group 1

Plot 5 represents the Decision Tree over Group 1 set but with the Predictors though X1-X7. This DT predicts with an accuracy of 78% (0.78).



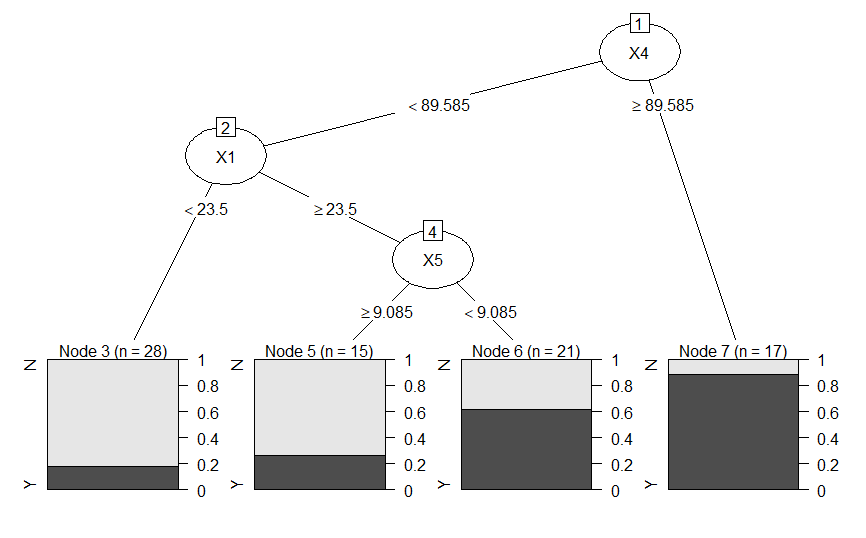
Plot 6 - DT over Y with Group 0

Plot 6 represents the Decision Tree over Group 0 set but with the Predictors though Y1-Y7. This DT predicts with an accuracy of 72% (0.72).



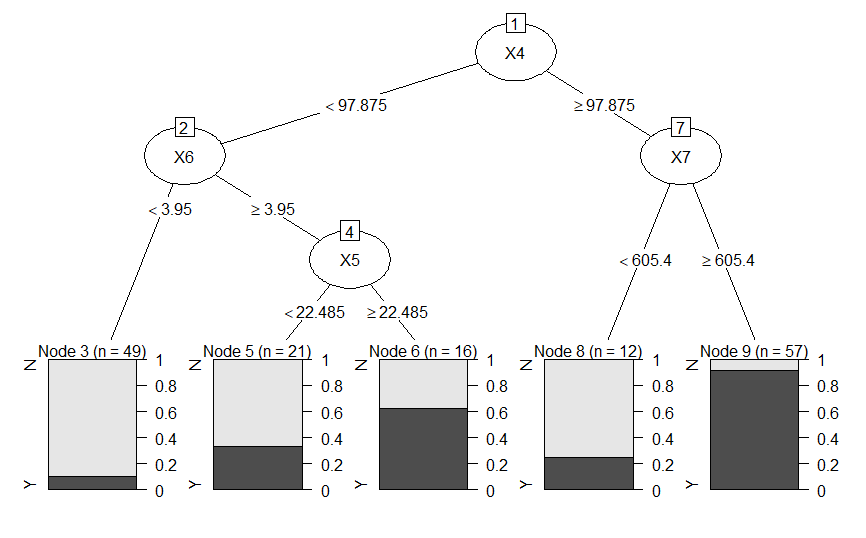
Plot 7 - DT over Y with Group 1

Plot 7 represents the Decision Tree over Group 1 set but with the Predictors though Y1-Y7. This DT predicts with an accuracy of 75% (0.75).



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 with an accuracy of 77% (0.77).



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 with an accuracy of 80% (0.8).

**Conclusion:**

The best Decision Tree generated is the DT generated over **Group 1 with the Predictors X1-X7 and Y1-Y7** as it has the highest **accuracy of 80%** - see **Plot 9** for the DT. The below is the Decision Tree’s summary, indicating the splits:

n= 155

node), split, n, loss, yval, (yprob)

\* denotes terminal node

1) root 155 77 N (0.5032258 0.4967742)

2) X4< 97.875 86 22 N (0.7441860 0.2558140)

4) X6< 3.95 49 5 N (0.8979592 0.1020408) \*

5) X6>=3.95 37 17 N (0.5405405 0.4594595)

10) X5< 22.485 21 7 N (0.6666667 0.3333333) \*

11) X5>=22.485 16 6 Y (0.3750000 0.6250000) \*

3) X4>=97.875 69 14 Y (0.2028986 0.7971014)

6) X7< 605.4 12 3 N (0.7500000 0.2500000) \*

7) X7>=605.4 57 5 Y (0.0877193 0.9122807) \*

Though the DT’s in Plot 1 and Plot 3 also have an accuracy of 80% - Plot 9 is better because it give a consistent accuracy of 80% when tested with different `rpart` configurations like minsplit, minbucket and maxdepth, more over it splits less frequently.

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

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