**Project Report**

Project: Prediction of losses Occurred to insurance company due to claims

Date:23-Apr-2016

Version: V2

Harshad Madhamshettiwar

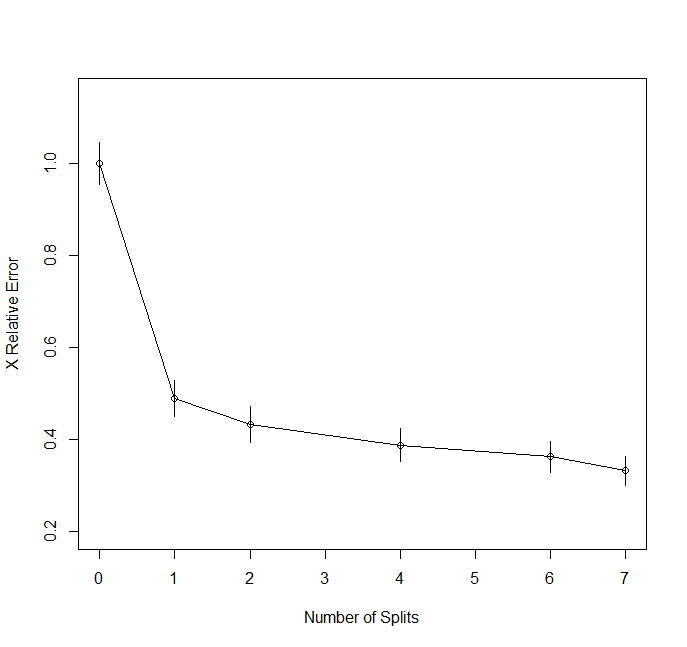
EPGP- BABD

2015-16

As decision tree with the predictive modeling is not adding lot of value to the busness case in hand. We are predicting the losses incured to the insurace comapany. Losses is continuous variable and the data also has the issue of outliers which we can not deal with without the intelligence from the Business. I have kept the scope of this extension of using decision tree for the project upto the selection of the variables only. This will serve as a confirmatory step towards variable selection donee using the linear regression (found in V1 of the report).

> insurance\_rpart <- rpart(hmformula, data = tree\_data.train,method = "anova",control = rpart.control(minsplit = 30,cp=0.01))

> plotcp(insurance\_rpart)



> rsq.rpart(insurance\_rpart)

Regression tree:

rpart(formula = hmformula, data = tree\_data.train, method = "anova",

control = rpart.control(minsplit = 30, cp = 0.01))

Variables actually used in tree construction:

[1] Age fuel\_dummy gender\_dummy married\_dummy Vehicle.Age

Root node error: 716497432/10706 = 66925

n= 10706

CP nsplit rel error xerror xstd

1 0.510929 0 1.00000 1.00010 0.046581

2 0.056509 1 0.48907 0.48921 0.038929

3 0.023298 2 0.43256 0.43278 0.039098

4 0.019479 4 0.38597 0.38695 0.035361

5 0.017818 6 0.34701 0.36144 0.033253

6 0.010000 7 0.32919 0.33078 0.032103

> print(insurance\_rpart)

n= 10677

node), split, n, deviance, yval

\* denotes terminal node

1) root 10677 695488500 389.8593

2) fuel\_dummy>=0.5 8144 161250900 286.7685

4) Age>=59.5 2923 57145110 190.7160 \*

5) Age< 59.5 5221 62039870 340.5440

10) Vehicle.Age>=10.5 2698 31969390 292.4166 \*

11) Vehicle.Age< 10.5 2523 17138580 392.0095 \*

3) fuel\_dummy< 0.5 2533 169406900 721.3127

6) Age>=25.5 1163 23956060 643.3457 \*

7) Age< 25.5 1370 132379600 787.4993

14) Vehicle.Age>=5.5 1091 43872990 729.3043 \*

15) Vehicle.Age< 5.5 279 70363460 1015.0650

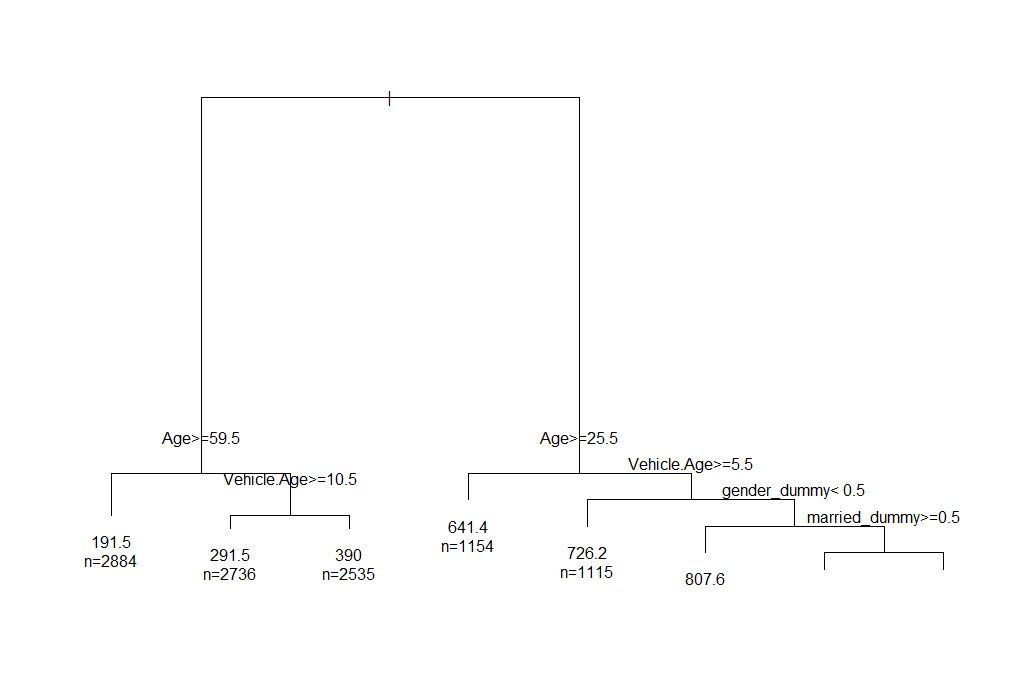
30) married\_dummy>=0.5 127 7993056 806.3228 \*

31) married\_dummy< 0.5 152 52213010 1189.4740

62) gender\_dummy< 0.5 83 6027765 896.0964 \*

63) gender\_dummy>=0.5 69 30448110 1542.3770 \*

> plot(insurance\_prune)



> plot(predict(insurance\_rpart),residuals(insurance\_rpart))