```
library(rpart)
library("rpart.plot")
library(ROCR)
```

STEP-1

```
data <- read.csv("C:/Users/hp/Downloads/HMEQ_Scrubbed (1)/HMEQ_Scrubbed.csv")
str(data)
```

```
## 'data.frame':    5960 obs. of  29 variables:
##  $ TARGET_BAD_FLAG    : int  1 1 1 1 0 1 1 1 1 1 ...
##  $ TARGET_LOSS_AMT    : int  641 1109 767 1425 0 335 1841 373 1217 1523 ...
##  $ LOAN               : int  1100 1300 1500 1500 1700 1700 1800 1800 2000 2000 ...
##  $ IMP_MORTDUE        : num  25860 70053 13500 65000 97800 ...
##  $ M_MORTDUE          : int  0 0 0 1 0 0 0 0 0 1 ...
##  $ IMP_VALUE          : num  39025 68400 16700 89000 112000 ...
##  $ M_VALUE            : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ IMP_YOJ            : num  10.5 7 4 7 3 9 5 11 3 16 ...
##  $ M_YOJ              : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ IMP_DEROG          : int  0 0 0 1 0 0 3 0 0 0 ...
##  $ M_DEROG            : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ IMP_DELINQ         : int  0 2 0 1 0 0 2 0 2 0 ...
##  $ M_DELINQ           : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ IMP_CLAGE          : num  94.4 121.8 149.5 174 93.3 ...
##  $ M_CLAGE            : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ IMP_NINQ           : int  1 0 1 1 0 1 1 0 1 0 ...
##  $ M_NINQ             : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ IMP_CLNO           : int  9 14 10 20 14 8 17 8 12 13 ...
##  $ M_CLNO             : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ IMP_DEBTINC        : num  35 35 35 35 35 ...
##  $ M_DEBTINC          : int  1 1 1 1 1 0 1 0 1 1 ...
##  $ FLAG.Job.Mgr       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FLAG.Job.Office    : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ FLAG.Job.Other     : int  1 1 1 0 0 1 1 1 1 0 ...
##  $ FLAG.Job.ProfExe   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FLAG.Job.Sales     : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ FLAG.Job.Self      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FLAG.Reason.DebtCon: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FLAG.Reason.HomeImp: int  1 1 1 0 1 1 1 1 1 1 ...
```

```
summary(data)
```

```
##  TARGET_BAD_FLAG  TARGET_LOSS_AMT      LOAN        IMP_MORTDUE
##  Min.   :0.0000   Min.   :    0   Min.   : 1100   Min.   :  2063
##  1st Qu.:0.0000   1st Qu.:    0   1st Qu.:11100   1st Qu.: 48139
##  Median :0.0000   Median :    0   Median :16300   Median : 65000
##  Mean   :0.1995   Mean   : 2676   Mean   :18608   Mean   : 72999
##  3rd Qu.:0.0000   3rd Qu.:    0   3rd Qu.:23300   3rd Qu.: 88200
##  Max.   :1.0000   Max.   :78987   Max.   :89900   Max.   :399550
##    M_MORTDUE         IMP_VALUE         M_VALUE           IMP_YOJ
```

```
##  Min.   :0.00000   Min.   :  8000   Min.   :0.00000   Min.   : 0.000
##  1st Qu.:0.00000   1st Qu.: 66490   1st Qu.:0.00000   1st Qu.: 3.000
##  Median :0.00000   Median : 89000   Median :0.00000   Median : 7.000
##  Mean   :0.08691   Mean   :101536   Mean   :0.01879   Mean   : 8.756
##  3rd Qu.:0.00000   3rd Qu.:119005   3rd Qu.:0.00000   3rd Qu.:12.000
##  Max.   :1.00000   Max.   :855909   Max.   :1.00000   Max.   :41.000
##      M_YOJ           IMP_DEROG         M_DEROG         IMP_DELINQ
##  Min.   :0.00000   Min.   : 0.0000   Min.   :0.0000   Min.   : 0.000
##  1st Qu.:0.00000   1st Qu.: 0.0000   1st Qu.:0.0000   1st Qu.: 0.000
##  Median :0.00000   Median : 0.0000   Median :0.0000   Median : 0.000
##  Mean   :0.08641   Mean   : 0.3431   Mean   :0.1188   Mean   : 0.503
##  3rd Qu.:0.00000   3rd Qu.: 0.0000   3rd Qu.:0.0000   3rd Qu.: 1.000
##  Max.   :1.00000   Max.   :10.0000   Max.   :1.0000   Max.   :15.000
##     M_DELINQ         IMP_CLAGE         M_CLAGE           IMP_NINQ
##  Min.   :0.00000   Min.   :   0.0   Min.   :0.00000   Min.   : 0.00
##  1st Qu.:0.00000   1st Qu.: 117.4   1st Qu.:0.00000   1st Qu.: 0.00
##  Median :0.00000   Median : 174.0   Median :0.00000   Median : 1.00
##  Mean   :0.09732   Mean   : 179.5   Mean   :0.05168   Mean   : 1.17
##  3rd Qu.:0.00000   3rd Qu.: 227.1   3rd Qu.:0.00000   3rd Qu.: 2.00
##  Max.   :1.00000   Max.   :1168.2   Max.   :1.00000   Max.   :17.00
##      M_NINQ           IMP_CLNO         M_CLNO          IMP_DEBTINC
##  Min.   :0.00000   Min.   : 0.00   Min.   :0.00000   Min.   :  0.5245
##  1st Qu.:0.00000   1st Qu.:15.00   1st Qu.:0.00000   1st Qu.: 30.7632
##  Median :0.00000   Median :20.00   Median :0.00000   Median : 35.0000
##  Mean   :0.08557   Mean   :21.25   Mean   :0.03725   Mean   : 34.0393
##  3rd Qu.:0.00000   3rd Qu.:26.00   3rd Qu.:0.00000   3rd Qu.: 37.9499
##  Max.   :1.00000   Max.   :71.00   Max.   :1.00000   Max.   :203.3122
##    M_DEBTINC       FLAG.Job.Mgr    FLAG.Job.Office  FLAG.Job.Other
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.2126   Mean   :0.1287   Mean   :0.1591   Mean   :0.4007
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  FLAG.Job.ProfExe FLAG.Job.Sales    FLAG.Job.Self     FLAG.Reason.DebtCon
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :1.0000
##  Mean   :0.2141   Mean   :0.01829   Mean   :0.03238   Mean   :0.6591
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.00000   Max.   :1.0000
##  FLAG.Reason.HomeImp
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.2987
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

```
head(data,6)
```

```
##   TARGET_BAD_FLAG TARGET_LOSS_AMT LOAN IMP_MORTDUE M_MORTDUE IMP_VALUE M_VALUE
## 1               1             641 1100       25860         0     39025       0
## 2               1            1109 1300       70053         0     68400       0
```

```
## 3              1             767 1500       13500         0     16700       0
## 4              1            1425 1500       65000         1     89000       1
## 5              0               0 1700       97800         0    112000       0
## 6              1             335 1700       30548         0     40320       0
##   IMP_YOJ M_YOJ IMP_DEROG M_DEROG IMP_DELINQ M_DELINQ IMP_CLAGE M_CLAGE
## 1    10.5     0         0       0          0        0  94.36667       0
## 2     7.0     0         0       0          2        0 121.83333       0
## 3     4.0     0         0       0          0        0 149.46667       0
## 4     7.0     1         1       1          1        1 174.00000       1
## 5     3.0     0         0       0          0        0  93.33333       0
## 6     9.0     0         0       0          0        0 101.46600       0
##   IMP_NINQ M_NINQ IMP_CLNO M_CLNO IMP_DEBTINC M_DEBTINC FLAG.Job.Mgr
## 1        1      0        9      0    35.00000         1            0
## 2        0      0       14      0    35.00000         1            0
## 3        1      0       10      0    35.00000         1            0
## 4        1      1       20      1    35.00000         1            0
## 5        0      0       14      0    35.00000         1            0
## 6        1      0        8      0    37.11361         0            0
##   FLAG.Job.Office FLAG.Job.Other FLAG.Job.ProfExe FLAG.Job.Sales FLAG.Job.Self
## 1               0              1                0              0             0
## 2               0              1                0              0             0
## 3               0              1                0              0             0
## 4               0              0                0              0             0
## 5               1              0                0              0             0
## 6               0              1                0              0             0
##   FLAG.Reason.DebtCon FLAG.Reason.HomeImp
## 1                   0                  1
## 2                   0                  1
## 3                   0                  1
## 4                   0                  0
## 5                   0                  1
## 6                   0                  1
```

```
dim(data)
```

```
## [1] 5960   29
```

STEP-2

```
SEED=1
set.seed(SEED)
```

```
data_flag= data
data_flag$TARGET_LOSS_AMT= NULL
head(data_flag)
```

```
##   TARGET_BAD_FLAG LOAN IMP_MORTDUE M_MORTDUE IMP_VALUE M_VALUE IMP_YOJ M_YOJ
## 1               1 1100       25860         0     39025       0    10.5     0
## 2               1 1300       70053         0     68400       0     7.0     0
## 3               1 1500       13500         0     16700       0     4.0     0
## 4               1 1500       65000         1     89000       1     7.0     1
```

```
## 5              0 1700       97800         0    112000       0     3.0     0
## 6              1 1700       30548         0     40320       0     9.0     0
##   IMP_DEROG M_DEROG IMP_DELINQ M_DELINQ IMP_CLAGE M_CLAGE IMP_NINQ M_NINQ
## 1         0       0          0        0  94.36667       0        1      0
## 2         0       0          2        0 121.83333       0        0      0
## 3         0       0          0        0 149.46667       0        1      0
## 4         1       1          1        1 174.00000       1        1      1
## 5         0       0          0        0  93.33333       0        0      0
## 6         0       0          0        0 101.46600       0        1      0
##   IMP_CLNO M_CLNO IMP_DEBTINC M_DEBTINC FLAG.Job.Mgr FLAG.Job.Office
## 1        9      0    35.00000         1            0               0
## 2       14      0    35.00000         1            0               0
## 3       10      0    35.00000         1            0               0
## 4       20      1    35.00000         1            0               0
## 5       14      0    35.00000         1            0               1
## 6        8      0    37.11361         0            0               0
##   FLAG.Job.Other FLAG.Job.ProfExe FLAG.Job.Sales FLAG.Job.Self
## 1              1                0              0             0
## 2              1                0              0             0
## 3              1                0              0             0
## 4              0                0              0             0
## 5              0                0              0             0
## 6              1                0              0             0
##   FLAG.Reason.DebtCon FLAG.Reason.HomeImp
## 1                   0                  1
## 2                   0                  1
## 3                   0                  1
## 4                   0                  0
## 5                   0                  1
## 6                   0                  1
```

```
FLAG= sample(c(TRUE,FALSE), nrow(data_flag), replace=TRUE, prob=c(0.7,0.3))
data_train= data_flag[FLAG, ]
data_test= data_flag[! FLAG, ]
```

```
dim(data_flag)
```

1

```
## [1] 5960   28
```

```
dim(data_train)
```

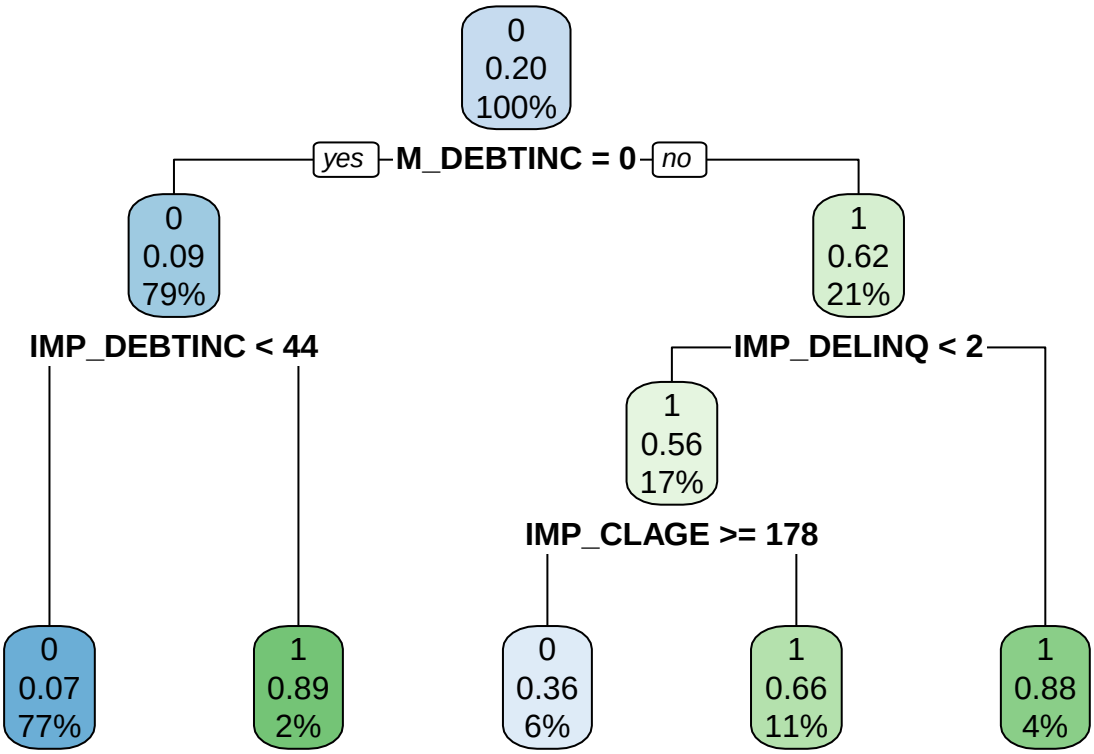
```
## [1] 4142   28
```

```
dim(data_test)
```

```
## [1] 1818   28
```

```
tr_set=rpart.control(maxdepth=10)
t1G = rpart(data=data_train, TARGET_BAD_FLAG ~ ., control = tr_set, method = "class", parms = list(spli
t1E = rpart(data=data_train, TARGET_BAD_FLAG ~ ., control = tr_set, method = "class", parms = list(spli
```

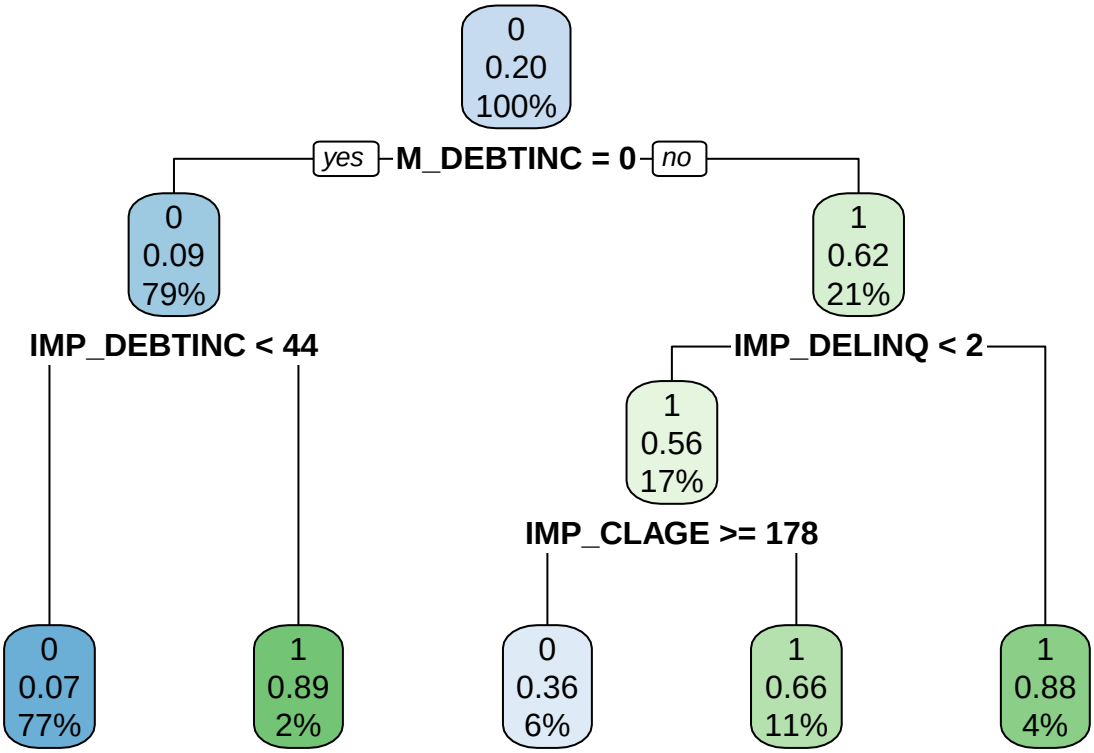
```
rpart.plot(t1G)
```




```
t1G$variable.importance
```

```
##   M_DEBTINC IMP_DEBTINC  IMP_DELINQ   IMP_CLAGE        LOAN     M_VALUE
##  399.453715   91.488049   37.425391   30.153485   18.763107   16.625183
##   IMP_VALUE IMP_MORTDUE    IMP_CLNO     IMP_YOJ
##    6.423116    4.559546    2.412279    2.050437
```

```
rpart.plot(t1E)
```

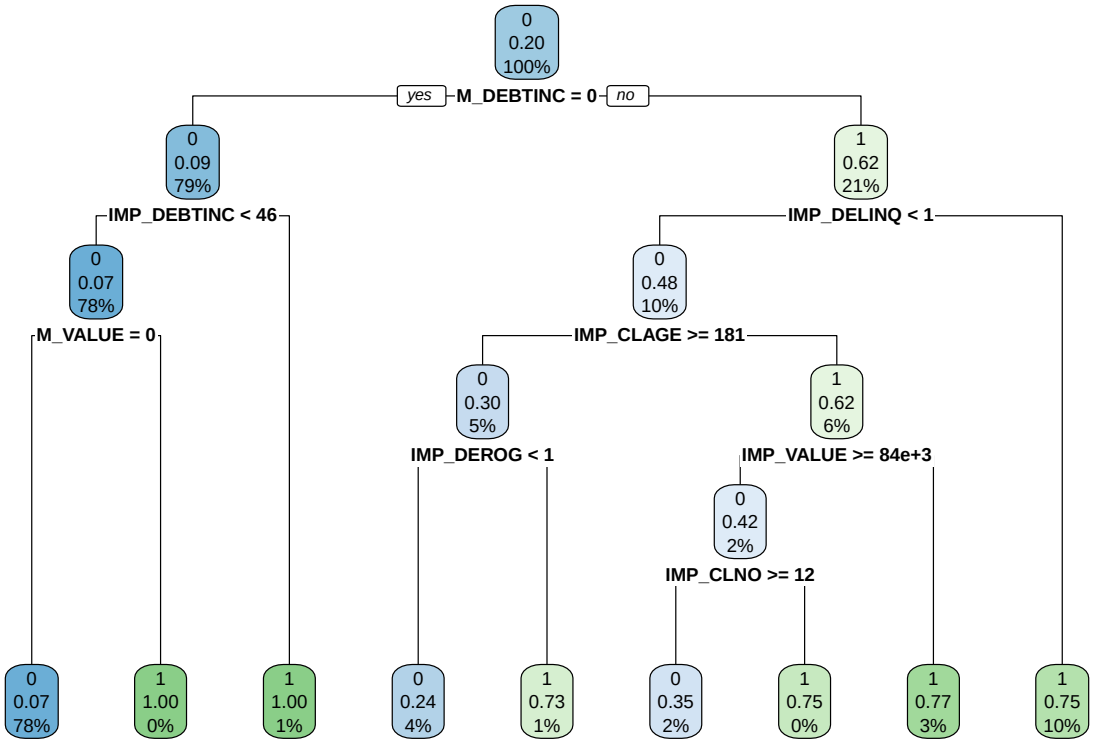



```
t1E$variable.importance
```

```
##   M_DEBTINC IMP_DEBTINC  IMP_DELINQ   IMP_CLAGE        LOAN     M_VALUE
##  533.397481  134.588883   46.494397   30.749923   24.521888   22.199895
##   IMP_VALUE IMP_MORTDUE    IMP_CLNO     IMP_YOJ
##    7.967967    5.783975    2.459994    2.090995
```

```
tr_set=rpart.control(maxdepth=10)
t1G1 = rpart(data=data_test, TARGET_BAD_FLAG ~ ., control = tr_set, method = "class", parms = list(spli
t1E1 = rpart(data=data_test, TARGET_BAD_FLAG ~ ., control = tr_set, method = "class", parms = list(spli
```

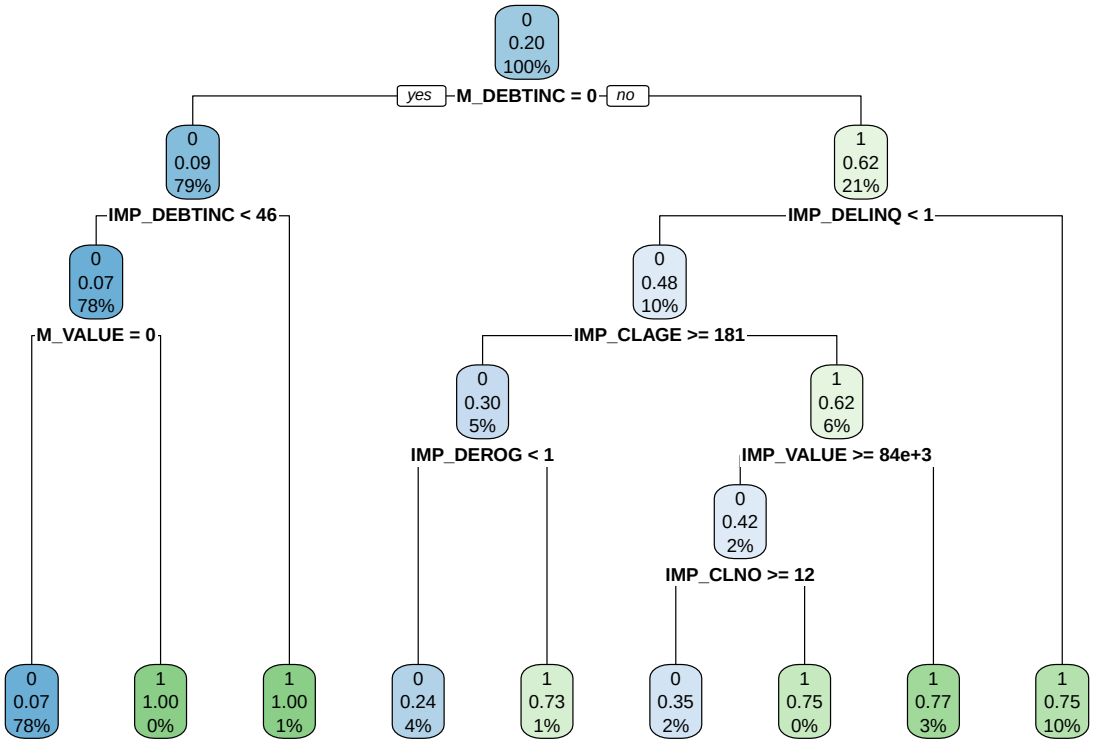
```
rpart.plot(t1G1)
```




```
t1G1$variable.importance
```

```
##        M_DEBTINC      IMP_DEBTINC       IMP_DELINQ          M_VALUE
##      170.5188649       39.1018260       21.3029284       18.0181645
##       IMP_DEROG             LOAN        IMP_CLAGE        IMP_VALUE
##       15.9253429       10.9918277        9.5981733        8.6264892
##         IMP_CLNO      IMP_MORTDUE           M_DEROG         M_DELINQ
##        8.4229578        7.1310676        3.3360009        2.7558268
##           M_NINQ FLAG.Job.ProfExe          IMP_YOJ   FLAG.Job.Other
##        2.3206963        1.5355987        1.3663664        1.1167991
##          M_CLAGE
##        0.5226727
```

```
rpart.plot(t1E1)
```




```
t1E1$variable.importance
```

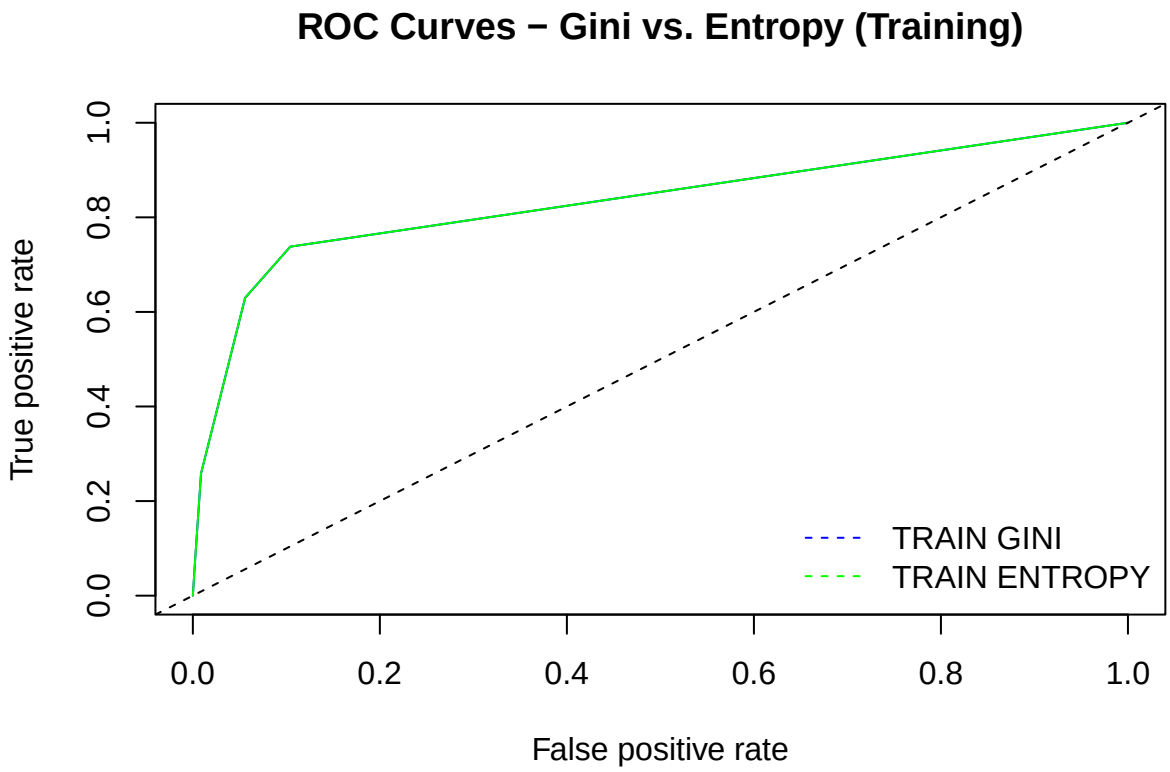
```
##        M_DEBTINC      IMP_DEBTINC          M_VALUE       IMP_DELINQ
##      229.1181956       58.6988395       26.6712227       24.9892440
##        IMP_DEROG             LOAN        IMP_CLAGE        IMP_VALUE
##       18.8528489       14.0629145        9.8085897        9.1488028
##         IMP_CLNO      IMP_MORTDUE          M_DEROG          M_DELINQ
##        8.9160029        8.0861916        3.5931350        2.9682420
##           M_NINQ FLAG.Job.ProfExe          IMP_YOJ   FLAG.Job.Other
##        2.4995722        1.6494196        1.3958921        1.1995779
##          M_CLAGE
##        0.5399878
```

####Training data

```
pG=predict(t1G, data_train, type="prob")
pG2 = prediction(pG[,2], data_train$TARGET_BAD_FLAG)
pG3 = performance(pG2, "tpr", "fpr")
```

```
pE= predict(t1E, data_train, type="prob")
pE2 = prediction(pE[,2], data_train$TARGET_BAD_FLAG)
pE3 = performance(pE2, "tpr", "fpr")
```

```
plot(pG3, col="blue", main = "ROC Curves - Gini vs. Entropy (Training)", lty = 1)
plot(pE3, col="green", add=TRUE, lty=1)
abline(0,1,lty=2)
legend("bottomright", c("TRAIN GINI","TRAIN ENTROPY"), col=c("blue","green"),bty="n",lty=2)
```




```
aucG = performance(pG2,"auc")@y.values
aucE = performance(pE2,"auc")@y.values
```

```
print(paste("TRAIN AUC GINI=", aucG))
```

```
## [1] "TRAIN AUC GINI= 0.83355126510574"
```

```
print(paste("TRAIN AUC ENTROPY=", aucE))
```

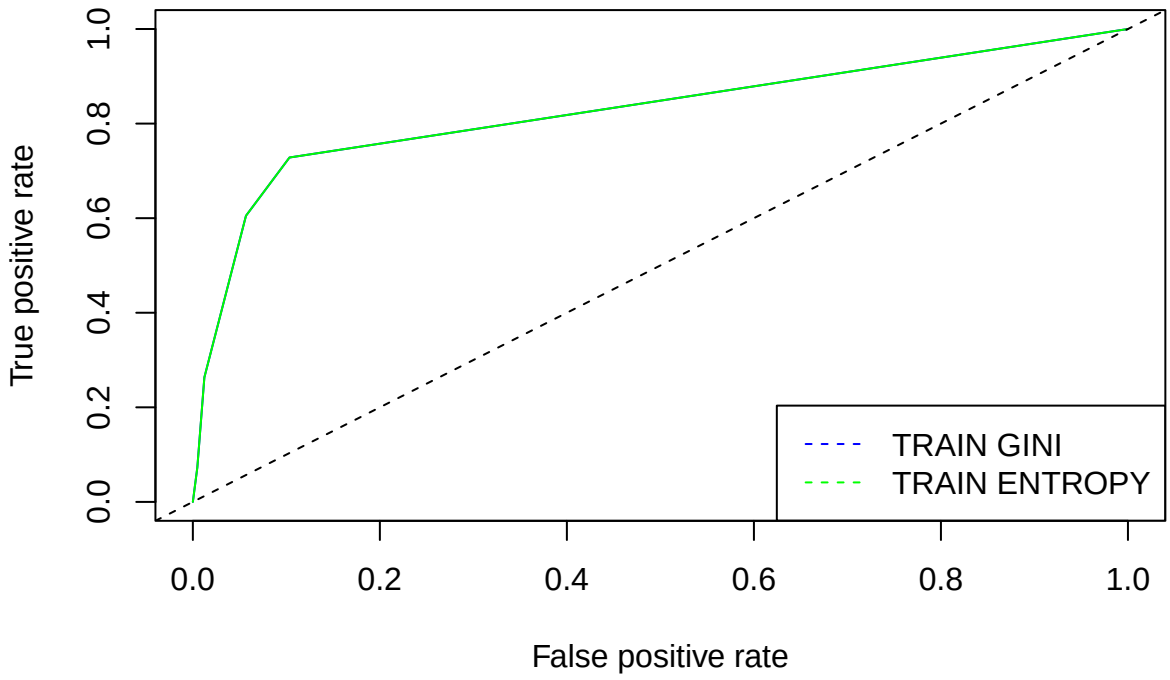
```
## [1] "TRAIN AUC ENTROPY= 0.83355126510574"
```

###Test data

```
pGT=predict(t1G, data_test)
pGT2 = prediction(pGT[,2], data_test$TARGET_BAD_FLAG)
pGT3 = performance(pGT2, "tpr", "fpr")
```

```
pET= predict(t1E, data_test)
pET2 = prediction(pET[,2], data_test$TARGET_BAD_FLAG)
pET3 = performance(pET2, "tpr", "fpr")
```

```
plot(pGT3, col="blue")
plot(pET3, col="green", add=TRUE)
abline(0,1,lty=2)
legend("bottomright", c("TRAIN GINI","TRAIN ENTROPY"), col=c("blue","green"),bty="y",lty=2)
```




```
aucG_T = performance(pGT2,"auc")@y.values
aucE_T = performance(pET2,"auc")@y.values
```

```
print(paste("TRAIN AUC GINI=", aucG_T))
```

```
## [1] "TRAIN AUC GINI= 0.826618121581281"
```

```
print(paste("TRAIN AUC ENTROPY=", aucE_T))
```

```
## [1] "TRAIN AUC ENTROPY= 0.826618121581281"
```

```
FLAG= sample(c(TRUE,FALSE), nrow(data_flag), replace=TRUE, prob=c(0.6,0.4))
data_train= data_flag[FLAG, ]
data_test= data_flag[! FLAG, ]
```

```
dim(data_flag)
```

2

```
## [1] 5960   28
```

```
dim(data_train)
```

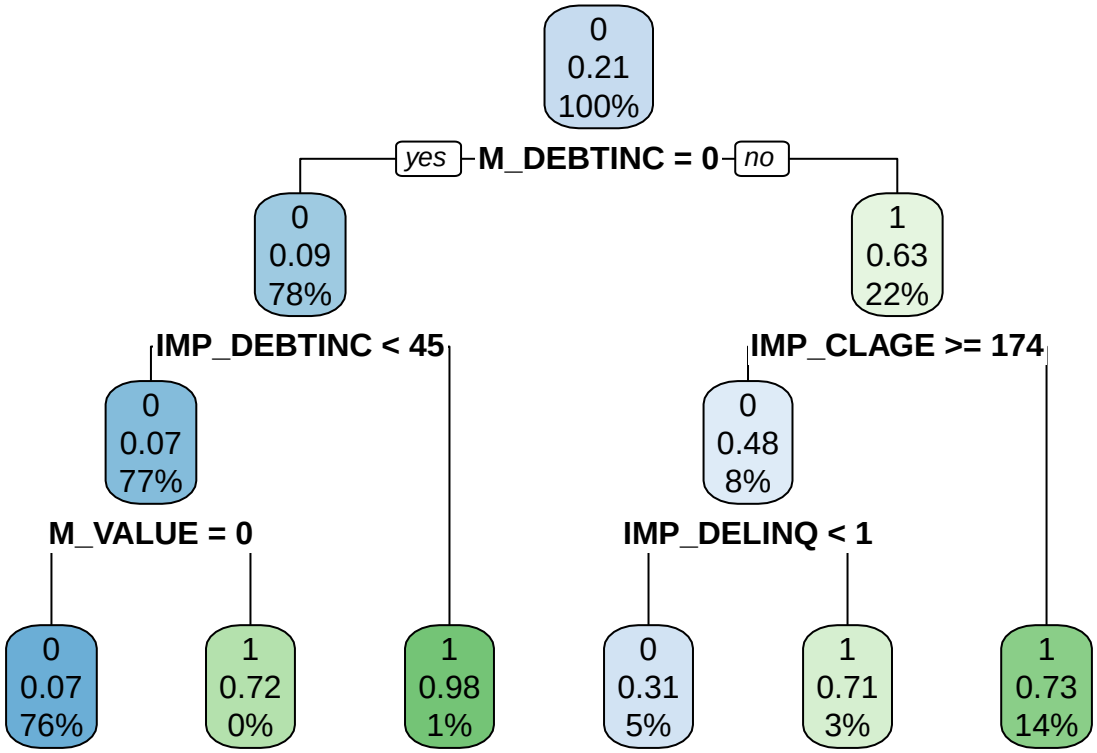
```
## [1] 3624   28
```

```
dim(data_test)
```

```
## [1] 2336   28
```

```
tr_set=rpart.control(maxdepth=10)
t1G = rpart(data=data_train, TARGET_BAD_FLAG ~ ., control = tr_set, method = "class", parms = list(spli
t1E = rpart(data=data_train, TARGET_BAD_FLAG ~ ., control = tr_set, method = "class", parms = list(spli
```

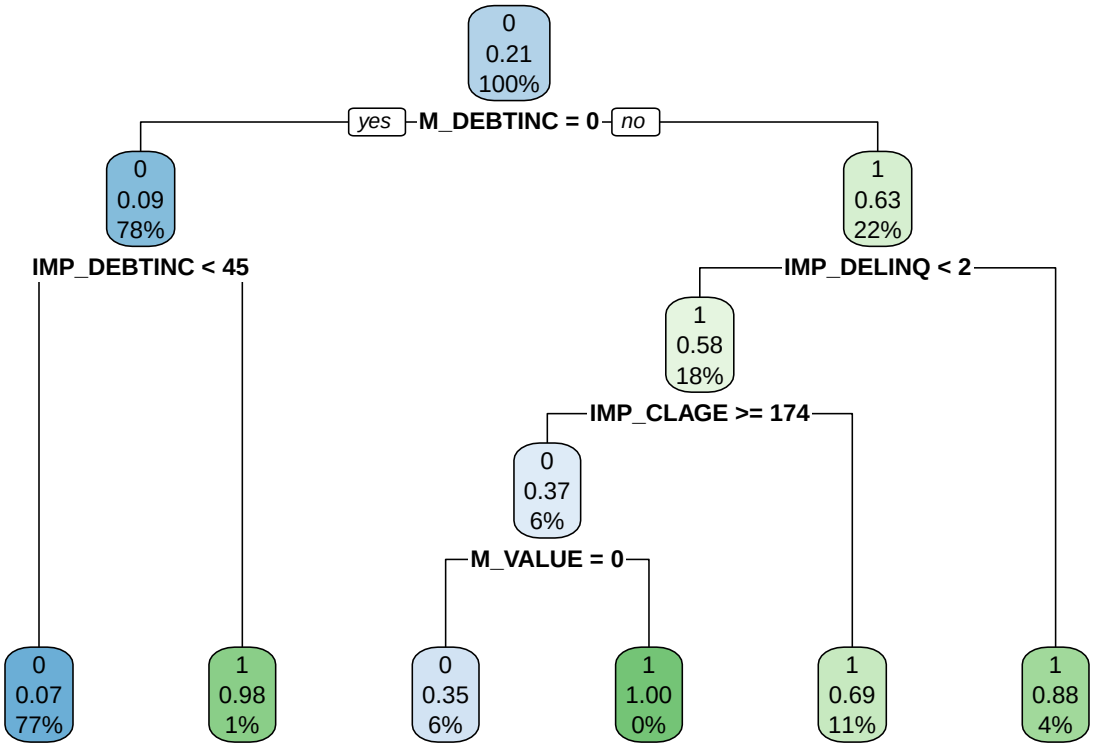
```
rpart.plot(t1G)
```




```
t1G$variable.importance
```

```
##       M_DEBTINC     IMP_DEBTINC      IMP_DELINQ         M_VALUE       IMP_CLAGE
##      372.693893       81.390521       32.311271       29.054745       23.359671
##            LOAN       IMP_VALUE       IMP_DEROG        IMP_CLNO         M_DEROG
##       20.811308        9.737657        8.003902        3.931996        3.394112
## FLAG.Job.Office        M_DELINQ     IMP_MORTDUE         IMP_YOJ
##        2.074179        1.885618        1.827457        1.589093
```

```
rpart.plot(t1E)
```

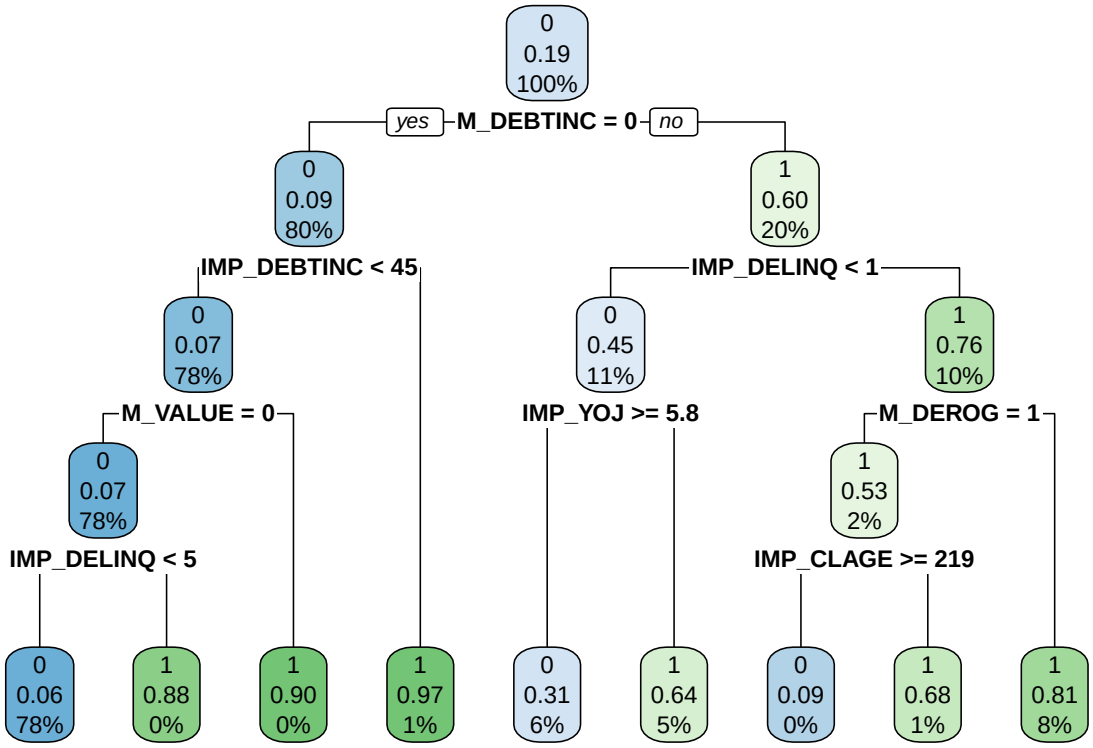



```
t1E$variable.importance
```

```
##   M_DEBTINC IMP_DEBTINC  IMP_DELINQ   IMP_CLAGE        LOAN     M_VALUE
##  492.854772  120.365251   39.012355   30.525064   26.229524   26.136976
##   IMP_VALUE    IMP_CLNO   IMP_DEROG IMP_MORTDUE     IMP_YOJ
##   13.485402    6.571410    4.350547    2.521636    1.592612
```

```
tr_set=rpart.control(maxdepth=10)
t1G1 = rpart(data=data_test, TARGET_BAD_FLAG ~ ., control = tr_set, method = "class", parms = list(spli
t1E1 = rpart(data=data_test, TARGET_BAD_FLAG ~ ., control = tr_set, method = "class", parms = list(spli
```

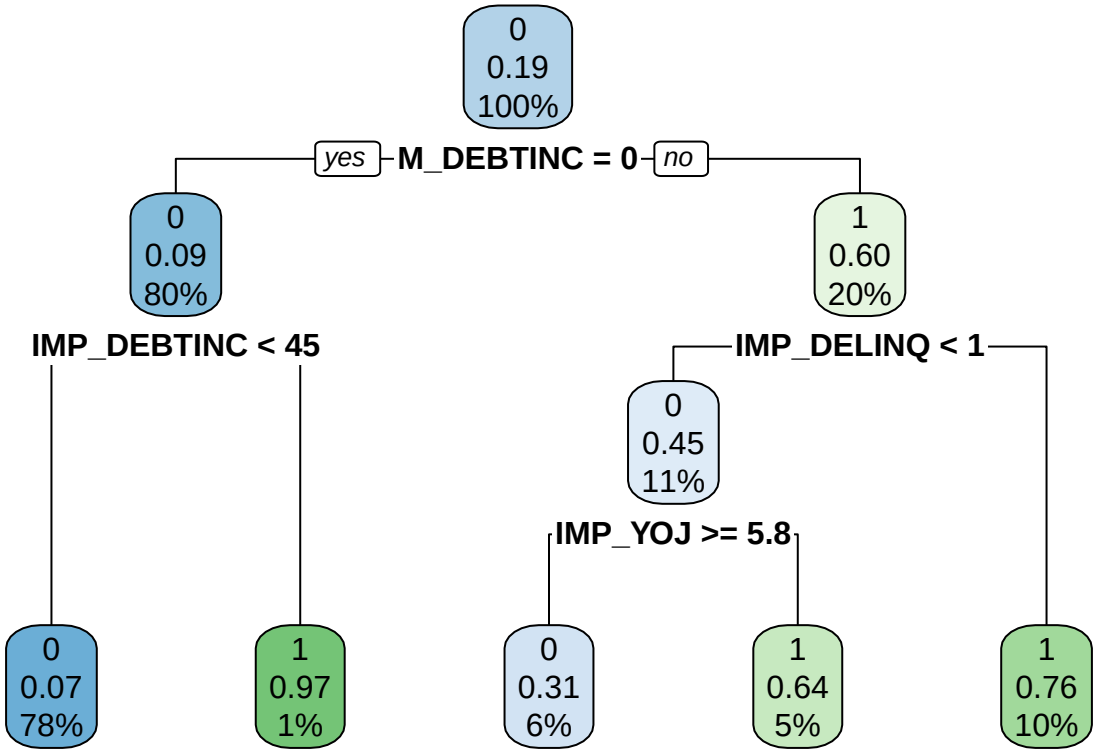
```
rpart.plot(t1G1)
```




```
t1G1$variable.importance
```

```
##   M_DEBTINC IMP_DEBTINC  IMP_DELINQ     M_VALUE   IMP_DEROG     IMP_YOJ
##  197.404181   50.752071   38.128128   24.926864   14.455027   13.689330
##     M_DEROG        LOAN   IMP_CLAGE    M_DELINQ      M_NINQ IMP_MORTDUE
##   10.078149    8.337994    7.056065    6.127645    5.806027    3.433385
##    IMP_CLNO     M_CLAGE      M_CLNO    IMP_NINQ
##    2.028830    1.894496    1.894496    1.233273
```

```
rpart.plot(t1E1)
```




```
t1E1$variable.importance
```

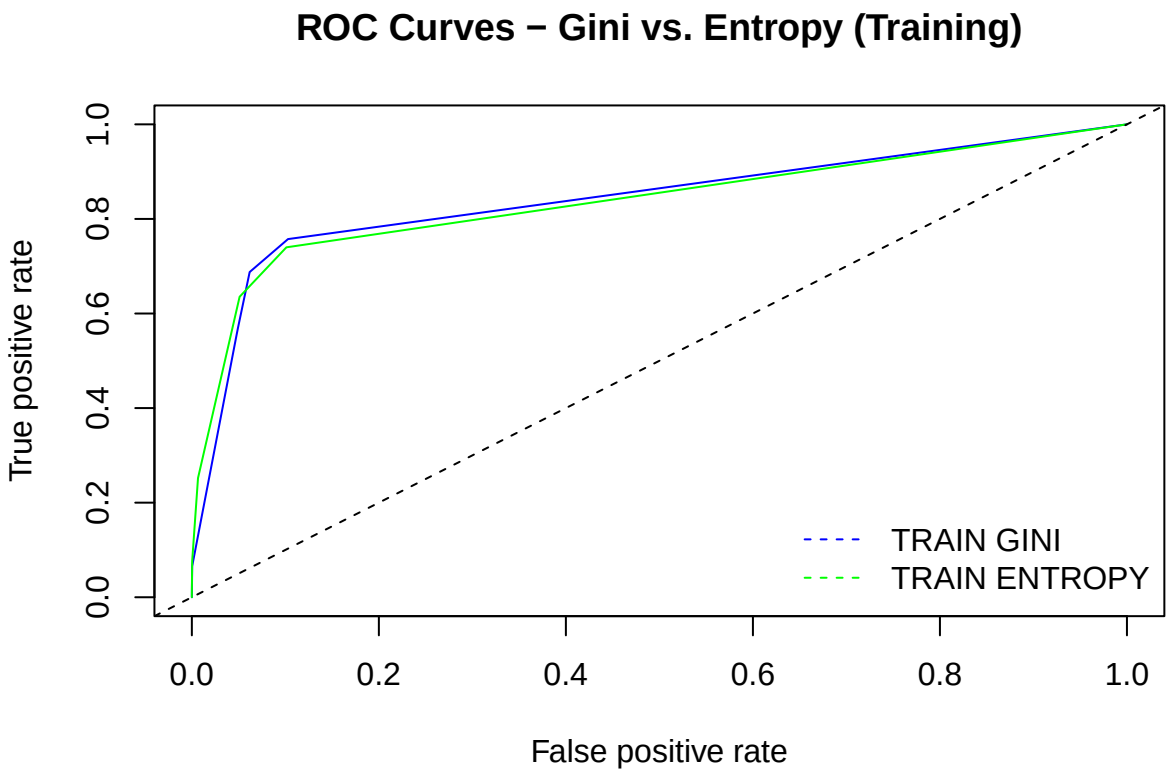
```
##   M_DEBTINC IMP_DEBTINC  IMP_DELINQ   IMP_DEROG     M_VALUE     IMP_YOJ
## 269.4342115  74.6632233  31.1585847  16.9907154  14.4338856  14.0209056
##        LOAN     M_DEROG IMP_MORTDUE    M_DELINQ      M_NINQ   IMP_CLAGE
##  10.4211391   4.6817379   3.9789862   3.4332745   3.2251972   1.3894591
##    IMP_NINQ    IMP_CLNO
##   1.2631446   0.7578868
```

####Training data

```
pG=predict(t1G, data_train, type="prob")
pG2 = prediction(pG[,2], data_train$TARGET_BAD_FLAG)
pG3 = performance(pG2, "tpr", "fpr")
```

```
pE= predict(t1E, data_train, type="prob")
pE2 = prediction(pE[,2], data_train$TARGET_BAD_FLAG)
pE3 = performance(pE2, "tpr", "fpr")
```

```
plot(pG3, col="blue", main = "ROC Curves - Gini vs. Entropy (Training)", lty = 1)
plot(pE3, col="green", add=TRUE, lty=1)
abline(0,1,lty=2)
legend("bottomright", c("TRAIN GINI","TRAIN ENTROPY"), col=c("blue","green"),bty="n",lty=2)
```




```
aucG = performance(pG2,"auc")@y.values
aucE = performance(pE2,"auc")@y.values
```

```
print(paste("TRAIN AUC GINI=", aucG))
```

```
## [1] "TRAIN AUC GINI= 0.841388959789249"
```

```
print(paste("TRAIN AUC ENTROPY=", aucE))
```

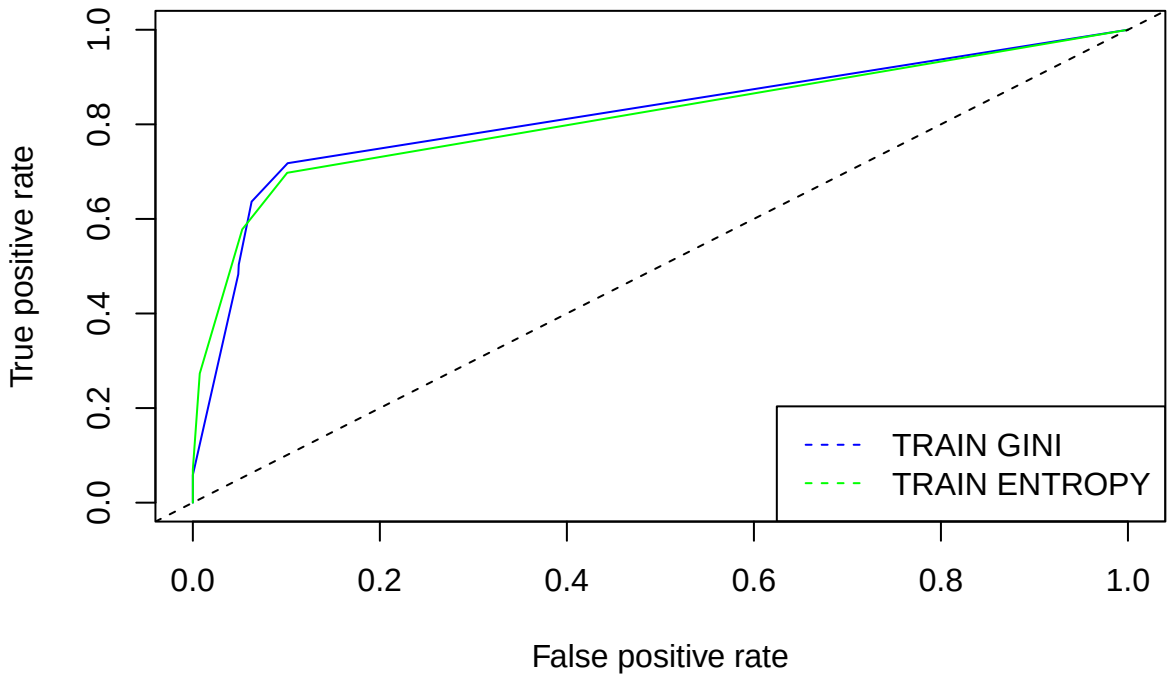
```
## [1] "TRAIN AUC ENTROPY= 0.837191684350355"
```

###Test data

```
pGT=predict(t1G, data_test)
pGT2 = prediction(pGT[,2], data_test$TARGET_BAD_FLAG)
pGT3 = performance(pGT2, "tpr", "fpr")
```

```
pET= predict(t1E, data_test)
pET2 = prediction(pET[,2], data_test$TARGET_BAD_FLAG)
pET3 = performance(pET2, "tpr", "fpr")
```

```
plot(pGT3, col="blue")
plot(pET3, col="green", add=TRUE)
abline(0,1,lty=2)
legend("bottomright", c("TRAIN GINI","TRAIN ENTROPY"), col=c("blue","green"),bty="y",lty=2)
```




```
aucG_T = performance(pGT2,"auc")@y.values
aucE_T = performance(pET2,"auc")@y.values
```

```
print(paste("TRAIN AUC GINI=", aucG_T))
```

```
## [1] "TRAIN AUC GINI= 0.819168637215165"
```

```
print(paste("TRAIN AUC ENTROPY=", aucE_T))
```

```
## [1] "TRAIN AUC ENTROPY= 0.814375523939332"
```

3

```
FLAG= sample(c(TRUE,FALSE), nrow(data_flag), replace=TRUE, prob=c(0.8,0.2))
data_train= data_flag[FLAG, ]
data_test= data_flag[! FLAG, ]
```

```
dim(data_flag)
```

```
## [1] 5960   28
```

```
dim(data_train)
```

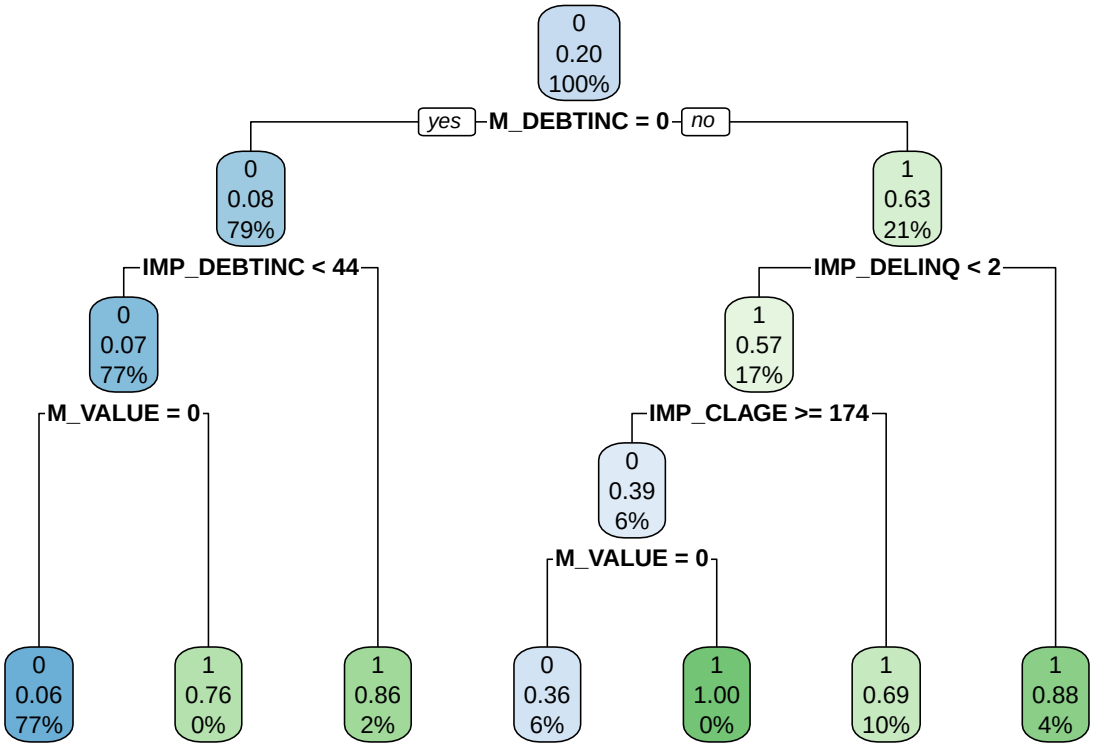
```
## [1] 4741   28
```

```
dim(data_test)
```

```
## [1] 1219   28
```

```
tr_set=rpart.control(maxdepth=10)
t1G = rpart(data=data_train, TARGET_BAD_FLAG ~ ., control = tr_set, method = "class", parms = list(spli
t1E = rpart(data=data_train, TARGET_BAD_FLAG ~ ., control = tr_set, method = "class", parms = list(spli
```

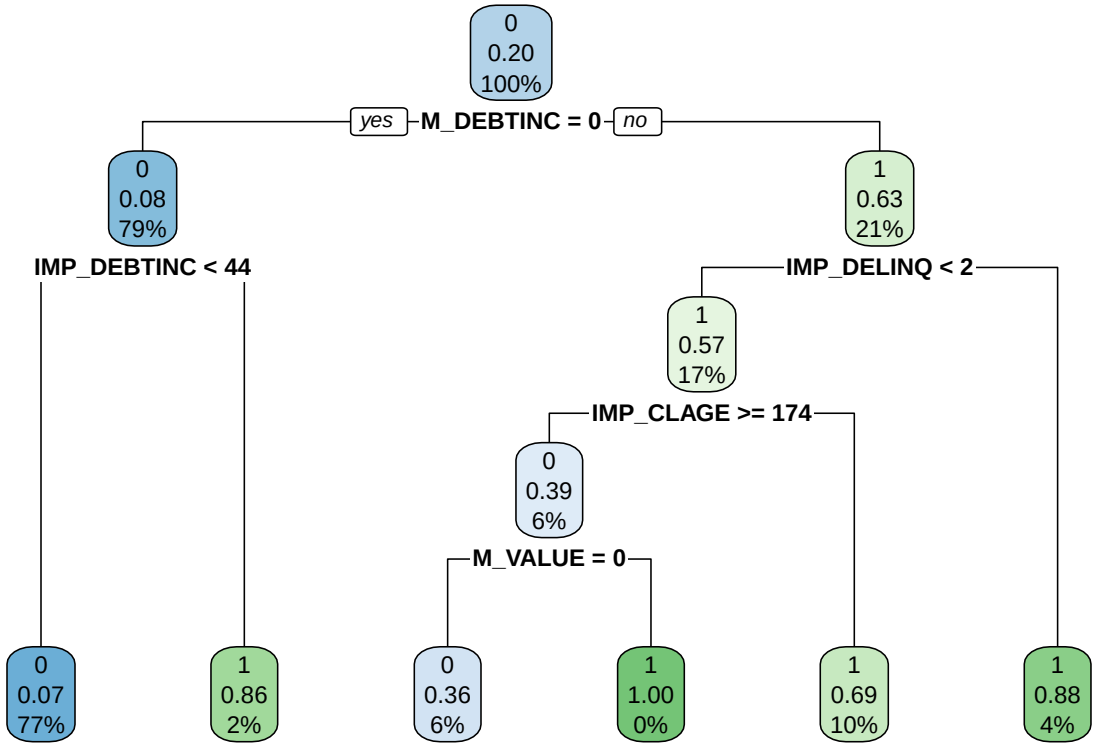
```
rpart.plot(t1G)
```




```
t1G$variable.importance
```

```
##   M_DEBTINC IMP_DEBTINC     M_VALUE  IMP_DELINQ   IMP_CLAGE        LOAN
## 476.5132340  97.6442898  45.3560096  40.9375325  34.7639816  21.2903231
##   IMP_VALUE   IMP_DEROG    IMP_CLNO     IMP_YOJ IMP_MORTDUE
##   8.2962739   6.7249852   3.4194080   3.0042908   0.6838816
```

```
rpart.plot(t1E)
```

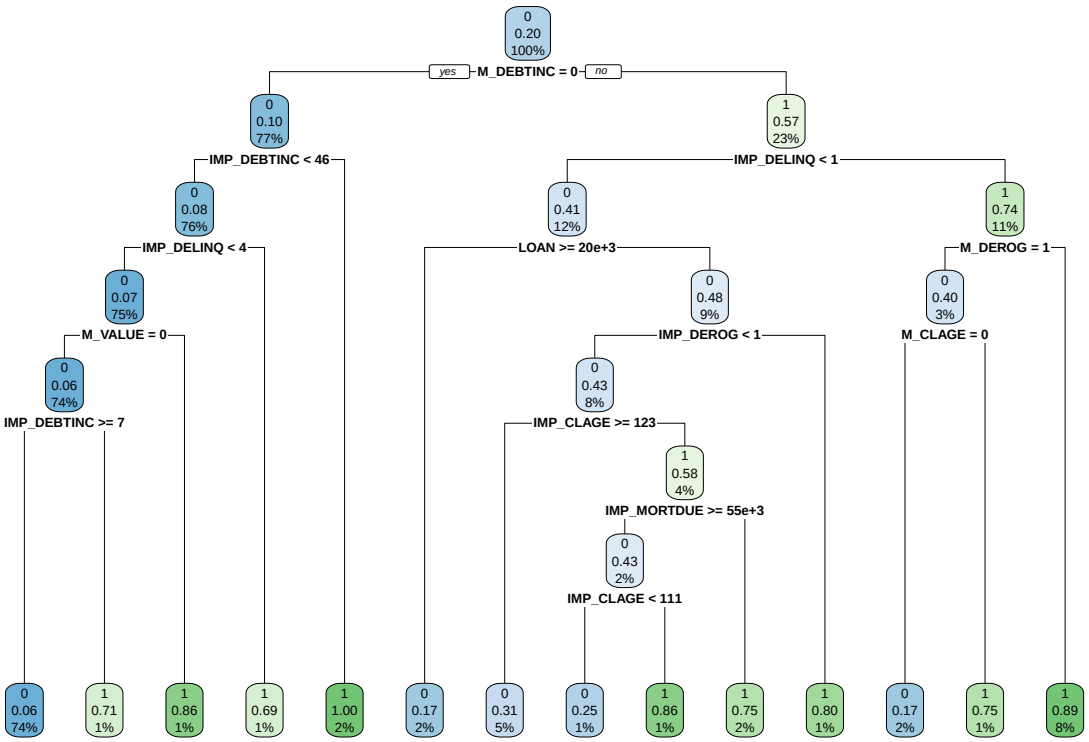



```
t1E$variable.importance
```

```
##   M_DEBTINC IMP_DEBTINC  IMP_DELINQ   IMP_CLAGE     M_VALUE        LOAN
## 637.5497526 145.5758905  51.9247518  35.7186873  32.6260435  27.9898889
##   IMP_VALUE   IMP_DEROG    IMP_CLNO     IMP_YOJ IMP_MORTDUE
##  10.4629748   8.9976780   3.5133135   3.1198674   0.7026627
```

```
tr_set=rpart.control(maxdepth=10)
t1G1 = rpart(data=data_test, TARGET_BAD_FLAG ~ ., control = tr_set, method = "class", parms = list(spli
t1E1 = rpart(data=data_test, TARGET_BAD_FLAG ~ ., control = tr_set, method = "class", parms = list(spli
```

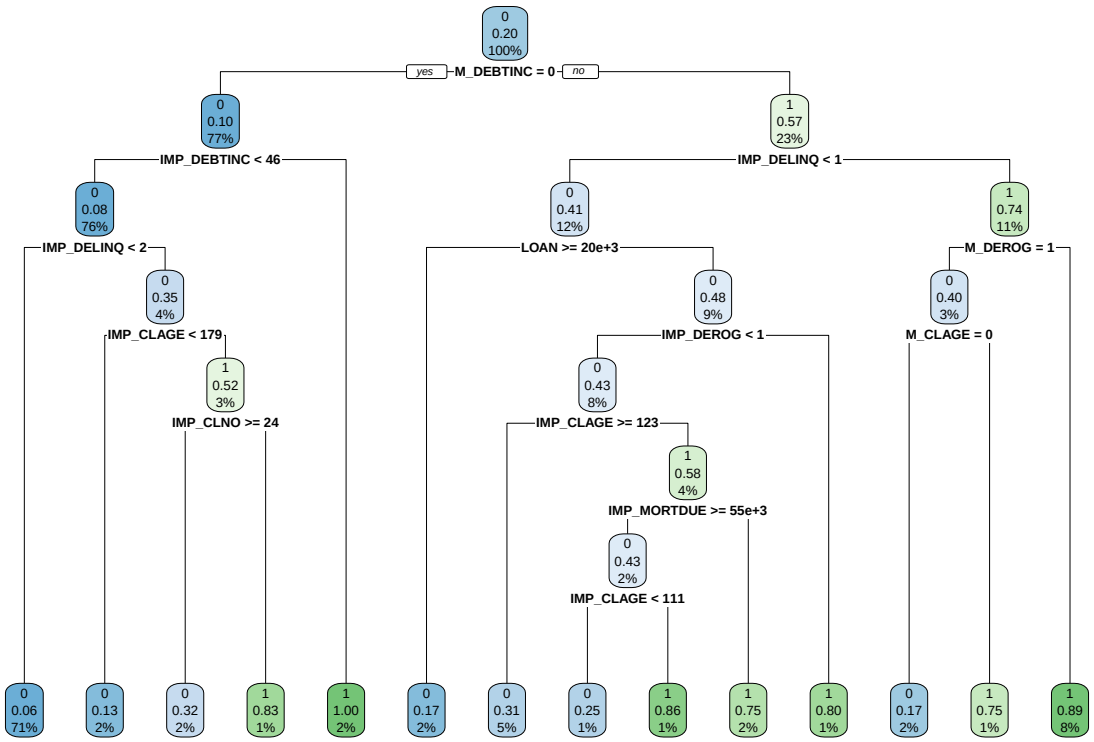
```
rpart.plot(t1G1)
```




```
t1G1$variable.importance
```

```
##        M_DEBTINC      IMP_DEBTINC       IMP_DELINQ          M_DEROG
##       95.4732937       39.2723654       28.3610287       17.9845226
##          M_VALUE           M_NINQ        IMP_DEROG        M_DELINQ
##       14.3150756       13.5664874       13.5597845       12.1425158
##          M_CLAGE           M_CLNO             LOAN       IMP_CLAGE
##       11.9168379       11.9168379       11.0615350        9.2105871
##         IMP_CLNO      IMP_MORTDUE        M_MORTDUE        IMP_VALUE
##        7.4225734        3.9729186        3.7240119        2.6131506
##          IMP_YOJ   FLAG.Job.Other   FLAG.Job.ProfExe         IMP_NINQ
##        1.8625955        0.8503539        0.8317487        0.2392920
##    FLAG.Job.Self
##        0.1530657
```

```
rpart.plot(t1E1)
```




```
t1E1$variable.importance
```

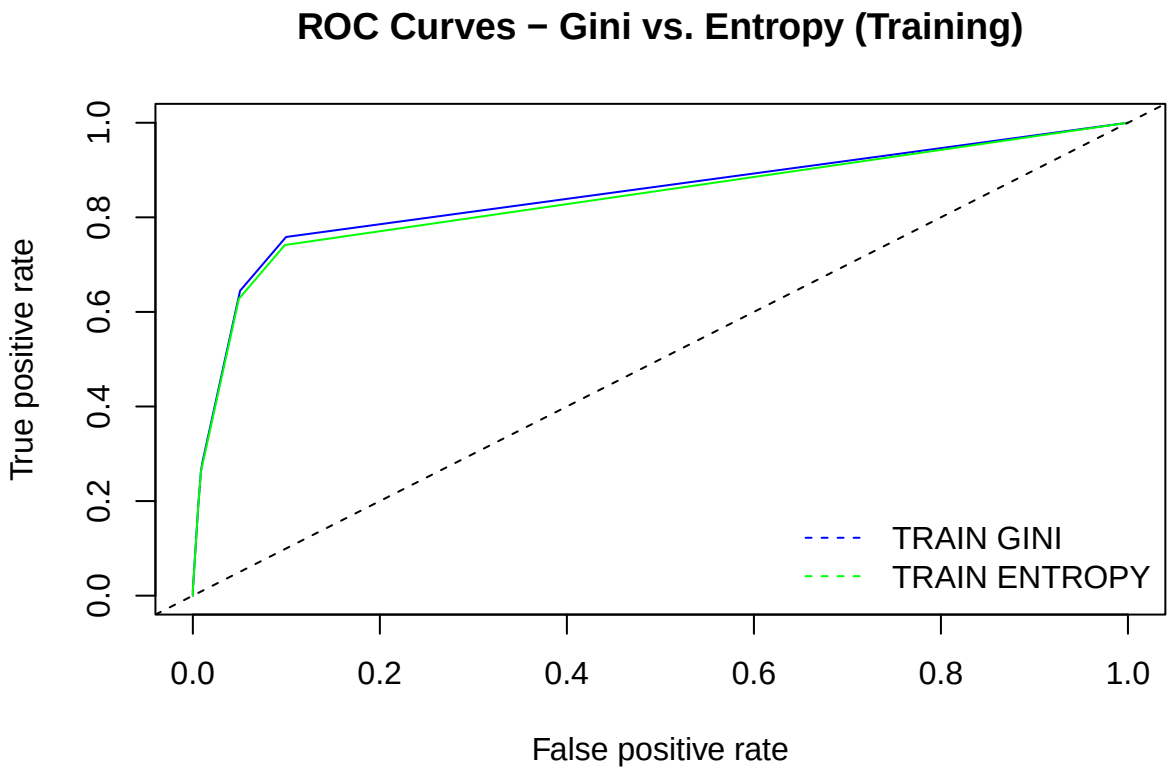
```
##        M_DEBTINC      IMP_DEBTINC       IMP_DELINQ          M_DEROG
##      127.7188256       49.7583717       38.4409256       21.4890298
##           M_NINQ        IMP_DEROG             LOAN        M_DELINQ
##       15.8181363       15.5101114       14.8967797       14.4409127
##        IMP_CLAGE          M_CLAGE           M_CLNO         IMP_CLNO
##       14.4276115       13.8130739       13.8130739       13.3303115
##          M_VALUE      IMP_MORTDUE        M_MORTDUE        IMP_VALUE
##        7.4309135        7.1104147        5.3703991        3.8787717
##          IMP_YOJ   FLAG.Job.Other FLAG.Job.ProfExe         IMP_NINQ
##        2.5698000        1.9498697        0.8900208        0.2522630
##    FLAG.Job.Self
##        0.1730550
```

####Training data

```
pG=predict(t1G, data_train, type="prob")
pG2 = prediction(pG[,2], data_train$TARGET_BAD_FLAG)
pG3 = performance(pG2, "tpr", "fpr")
```

```
pE= predict(t1E, data_train, type="prob")
pE2 = prediction(pE[,2], data_train$TARGET_BAD_FLAG)
pE3 = performance(pE2, "tpr", "fpr")
```

```
plot(pG3, col="blue", main = "ROC Curves - Gini vs. Entropy (Training)", lty = 1)
plot(pE3, col="green", add=TRUE, lty=1)
abline(0,1,lty=2)
legend("bottomright", c("TRAIN GINI","TRAIN ENTROPY"), col=c("blue","green"),bty="n",lty=2)
```




```
aucG = performance(pG2,"auc")@y.values
aucE = performance(pE2,"auc")@y.values
```

```
print(paste("TRAIN AUC GINI=", aucG))
```

```
## [1] "TRAIN AUC GINI= 0.846426332376138"
```

```
print(paste("TRAIN AUC ENTROPY=", aucE))
```

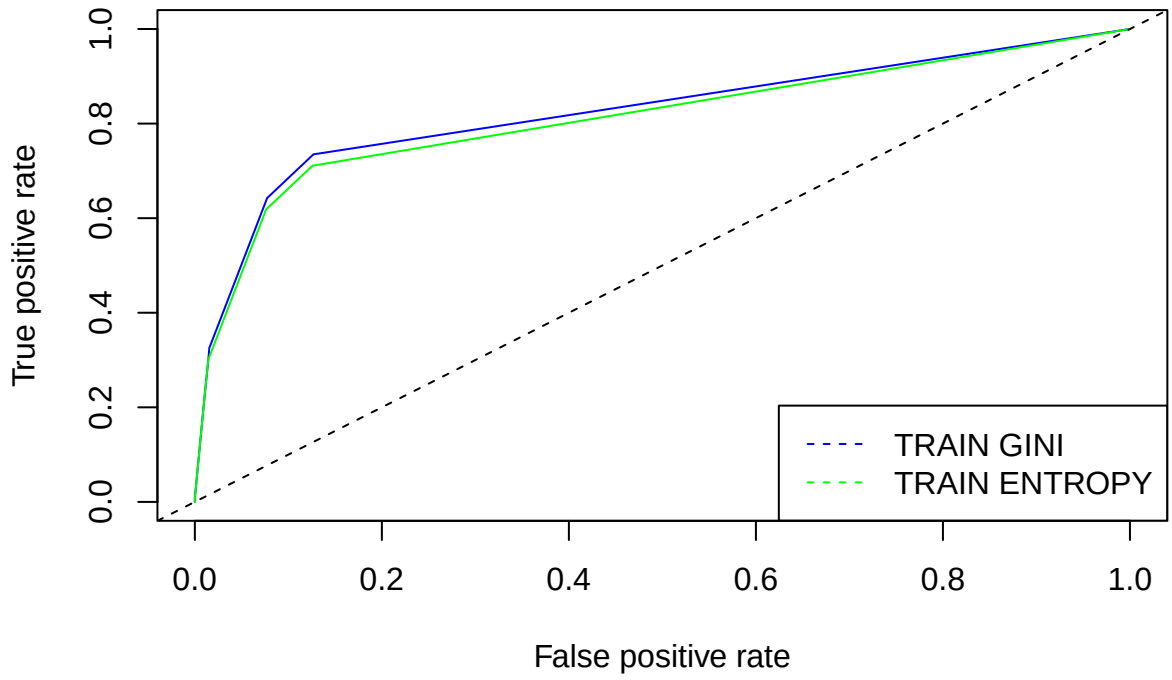
```
## [1] "TRAIN AUC ENTROPY= 0.838025827469815"
```

###Test data

```
pGT=predict(t1G, data_test)
pGT2 = prediction(pGT[,2], data_test$TARGET_BAD_FLAG)
pGT3 = performance(pGT2, "tpr", "fpr")
```

```
pET= predict(t1E, data_test)
pET2 = prediction(pET[,2], data_test$TARGET_BAD_FLAG)
pET3 = performance(pET2, "tpr", "fpr")
```

```
plot(pGT3, col="blue")
plot(pET3, col="green", add=TRUE)
abline(0,1,lty=2)
legend("bottomright", c("TRAIN GINI","TRAIN ENTROPY"), col=c("blue","green"),bty="y",lty=2)
```




```
aucG_T = performance(pGT2,"auc")@y.values
aucE_T = performance(pET2,"auc")@y.values
```

```
print(paste("TRAIN AUC GINI=", aucG_T))
```

```
## [1] "TRAIN AUC GINI= 0.824067403635159"
```

```
print(paste("TRAIN AUC ENTROPY=", aucE_T))
```

```
## [1] "TRAIN AUC ENTROPY= 0.811423011634166"
```

The ROC curves for both trees are optimal According to the results of my code 'Gini' is slightly better than 'Entropy'

```
data_amt=data
data_amt$TARGET_BAD_FLAG = NULL
head(data_amt)
```

STep-3

```
##   TARGET_LOSS_AMT LOAN IMP_MORTDUE M_MORTDUE IMP_VALUE M_VALUE IMP_YOJ M_YOJ
## 1             641 1100       25860         0     39025       0    10.5     0
## 2            1109 1300       70053         0     68400       0     7.0     0
## 3             767 1500       13500         0     16700       0     4.0     0
## 4            1425 1500       65000         1     89000       1     7.0     1
## 5               0 1700       97800         0    112000       0     3.0     0
## 6             335 1700       30548         0     40320       0     9.0     0
##   IMP_DEROG M_DEROG IMP_DELINQ M_DELINQ IMP_CLAGE M_CLAGE IMP_NINQ M_NINQ
## 1         0       0          0        0  94.36667       0        1      0
## 2         0       0          2        0 121.83333       0        0      0
## 3         0       0          0        0 149.46667       0        1      0
## 4         1       1          1        1 174.00000       1        1      1
## 5         0       0          0        0  93.33333       0        0      0
## 6         0       0          0        0 101.46600       0        1      0
##   IMP_CLNO M_CLNO IMP_DEBTINC M_DEBTINC FLAG.Job.Mgr FLAG.Job.Office
## 1        9      0    35.00000         1            0               0
## 2       14      0    35.00000         1            0               0
## 3       10      0    35.00000         1            0               0
## 4       20      1    35.00000         1            0               0
## 5       14      0    35.00000         1            0               1
## 6        8      0    37.11361         0            0               0
##   FLAG.Job.Other FLAG.Job.ProfExe FLAG.Job.Sales FLAG.Job.Self
## 1              1                0              0             0
## 2              1                0              0             0
## 3              1                0              0             0
## 4              0                0              0             0
## 5              0                0              0             0
## 6              1                0              0             0
##   FLAG.Reason.DebtCon FLAG.Reason.HomeImp
## 1                   0                  1
## 2                   0                  1
## 3                   0                  1
## 4                   0                  0
## 5                   0                  1
## 6                   0                  1
```

```
FLAG=sample(c(TRUE,FALSE), nrow(data_amt), replace=TRUE, prob=c(0.7,0.3))
data_train_s3=data_amt[FLAG, ]
data_test_s3=data_amt[! FLAG, ]
```

```
mean(data_amt$TARGET_LOSS_AMT)
```

1

```
## [1] 2676.163
```

```
mean(data_train_s3$TARGET_LOSS_AMT)
```

```
## [1] 2672.78
```

```
mean(data_test_s3$TARGET_LOSS_AMT)
```

```
## [1] 2683.926
```

```
dim(data_amt)
```

```
## [1] 5960   28
```

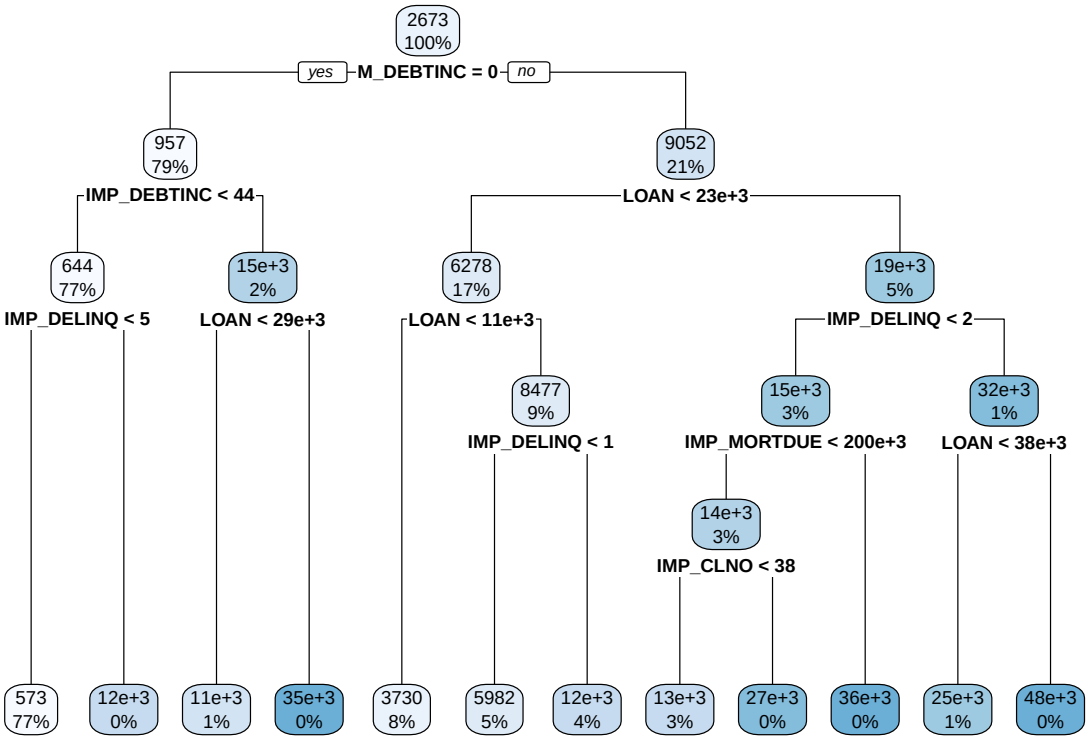
```
dim(data_train_s3)
```

```
## [1] 4151   28
```

```
dim(data_test_s3)
```

```
## [1] 1809   28
```

```
T1A=rpart(data=data_train_s3, TARGET_LOSS_AMT~., control=tr_set, method="anova")
rpart.plot(T1A)
```

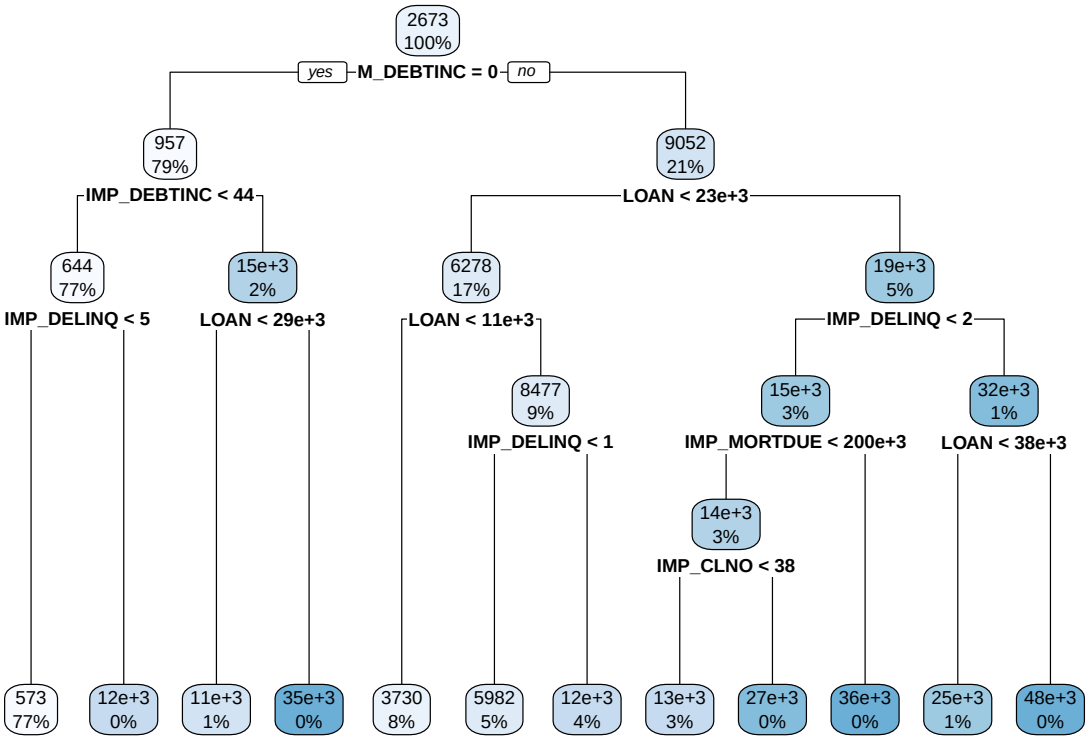


training

```
T1A$variable.importance
```

```
##          M_DEBTINC               LOAN          IMP_DELINQ         IMP_DEBTINC
##        45448520886        41668881373        16612572589        14641506027
##         IMP_VALUE        IMP_MORTDUE           IMP_CLNO           IMP_DEROG
##         8013498059         6991092645         3234480209         2005176742
##           M_VALUE            IMP_YOJ FLAG.Reason.HomeImp FLAG.Reason.DebtCon
##         1652673487         1341986369         1076618072         1040327575
##           M_DEROG           M_DELINQ              M_NINQ              M_CLNO
##          868963720          687087593          586045299          404169172
##         IMP_CLAGE      FLAG.Job.Self
##          399195465          134185607
```

```
T1P=rpart(data=data_train_s3, TARGET_LOSS_AMT~., control=tr_set, method="anova")
rpart.plot(T1P)
```




```
T1P$variable.importance
```

```
##          M_DEBTINC               LOAN          IMP_DELINQ         IMP_DEBTINC
##        45448520886        41668881373        16612572589        14641506027
##          IMP_VALUE         IMP_MORTDUE            IMP_CLNO           IMP_DEROG
##         8013498059         6991092645         3234480209         2005176742
##            M_VALUE            IMP_YOJ FLAG.Reason.HomeImp FLAG.Reason.DebtCon
##         1652673487         1341986369         1076618072         1040327575
##            M_DEROG           M_DELINQ              M_NINQ              M_CLNO
##          868963720          687087593          586045299          404169172
##          IMP_CLAGE      FLAG.Job.Self
##          399195465          134185607
```

```
P1A=predict(T1A,data_test_s3)
RMSE1a=sqrt(mean((data_test_s3$TARGET_LOSS_AMT-P1A)^2))

P1P=predict(T1P,data_test_s3)
RMSE1p=sqrt(mean((data_test_s3$TARGET_LOSS_AMT-P1P)^2))
```

```
print(paste("TEST RMSE ANOVA =", RMSE1a))
```

Test

```
## [1] "TEST RMSE ANOVA = 5502.02184703786"
```

```
print(paste("TEST RMSE POISSON =", RMSE1p))
```

```
## [1] "TEST RMSE POISSON = 5502.02184703786"
```

2

```
FLAG=sample(c(TRUE,FALSE), nrow(data_amt), replace=TRUE, prob=c(0.6,0.4))
data_train_s3=data_amt[FLAG, ]
data_test_s3=data_amt[! FLAG, ]
```

```
mean(data_amt$TARGET_LOSS_AMT)
```

```
## [1] 2676.163
```

```
mean(data_train_s3$TARGET_LOSS_AMT)
```

```
## [1] 2611.156
```

```
mean(data_test_s3$TARGET_LOSS_AMT)
```

```
## [1] 2775.326
```

```
dim(data_amt)
```

```
## [1] 5960   28
```

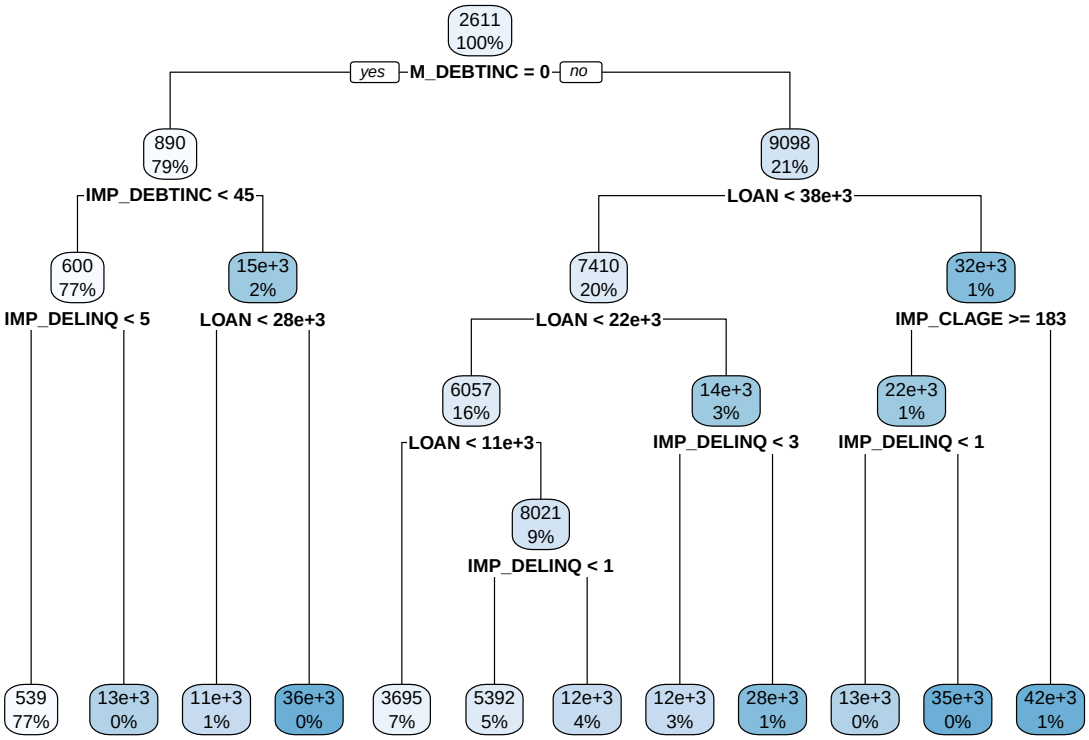
```
dim(data_train_s3)
```

```
## [1] 3600   28
```

```
dim(data_test_s3)
```

```
## [1] 2360   28
```

```
T1A=rpart(data=data_train_s3, TARGET_LOSS_AMT~., control=tr_set, method="anova")
rpart.plot(T1A)
```

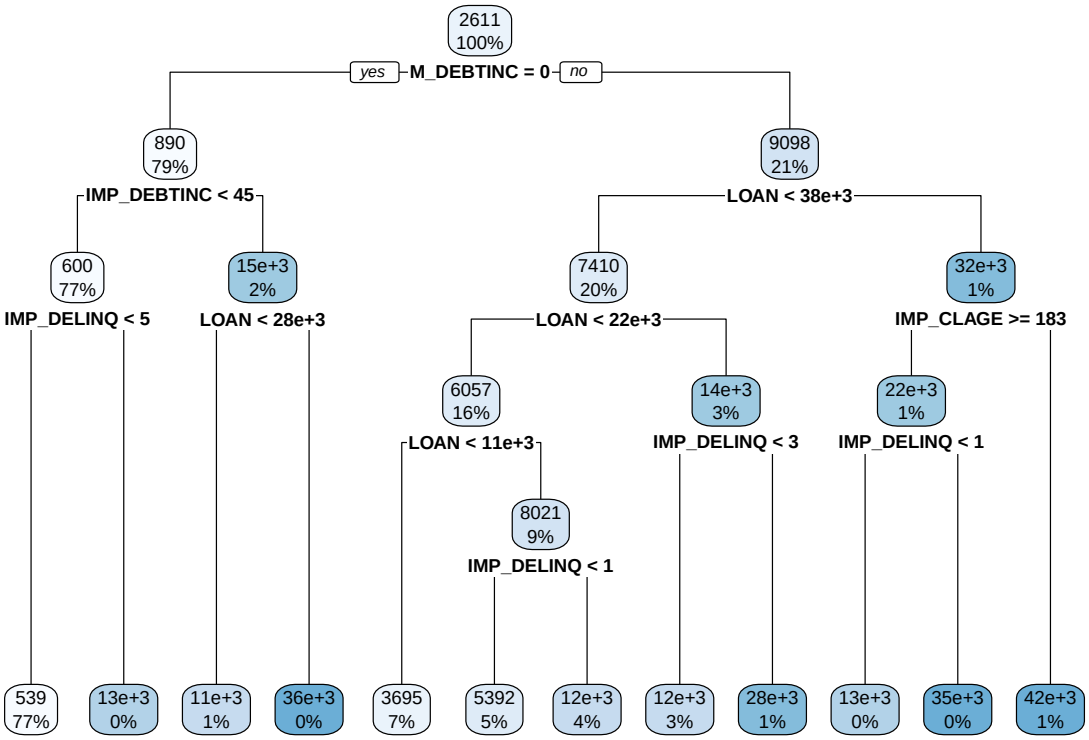


training

```
T1A$variable.importance
```

```
##                LOAN           M_DEBTINC           IMP_DELINQ         IMP_DEBTINC
##          47507682967         40199462013         15081603217         11720522943
##           IMP_CLAGE           IMP_VALUE         IMP_MORTDUE           IMP_DEROG
##           7083713086          5423410394          4364285473          3130235966
##             IMP_YOJ             M_VALUE             IMP_NINQ             M_DEROG
##           2702037973          2203404205           997735898           721405276
## FLAG.Reason.HomeImp            M_DELINQ FLAG.Reason.DebtCon              M_NINQ
##            613188257           597025056           541649627           422892748
##             IMP_CLNO
##            273636484
```

```
T1P=rpart(data=data_train_s3, TARGET_LOSS_AMT~., control=tr_set, method="anova")
rpart.plot(T1P)
```




```
T1P$variable.importance
```

```
##                LOAN           M_DEBTINC           IMP_DELINQ         IMP_DEBTINC
##          47507682967         40199462013         15081603217         11720522943
##           IMP_CLAGE           IMP_VALUE         IMP_MORTDUE           IMP_DEROG
##           7083713086          5423410394          4364285473          3130235966
##             IMP_YOJ             M_VALUE             IMP_NINQ             M_DEROG
##           2702037973          2203404205           997735898           721405276
## FLAG.Reason.HomeImp            M_DELINQ FLAG.Reason.DebtCon              M_NINQ
##            613188257           597025056           541649627           422892748
##             IMP_CLNO
##            273636484
```

```
P1A=predict(T1A,data_test_s3)
RMSE1a=sqrt(mean((data_test_s3$TARGET_LOSS_AMT-P1A)^2))

P1P=predict(T1P,data_test_s3)
RMSE1p=sqrt(mean((data_test_s3$TARGET_LOSS_AMT-P1P)^2))
```

```
print(paste("TEST RMSE ANOVA =", RMSE1a))
```

Test

```
## [1] "TEST RMSE ANOVA = 5480.9754217008"
```

```
print(paste("TEST RMSE POISSON =", RMSE1p))
```

```
## [1] "TEST RMSE POISSON = 5480.9754217008"
```

```
FLAG=sample(c(TRUE,FALSE), nrow(data_amt), replace=TRUE, prob=c(0.8,0.2))
data_train_s3=data_amt[FLAG, ]
data_test_s3=data_amt[! FLAG, ]
```

```
mean(data_amt$TARGET_LOSS_AMT)
```

3

```
## [1] 2676.163
```

```
mean(data_train_s3$TARGET_LOSS_AMT)
```

```
## [1] 2631.277
```

```
mean(data_test_s3$TARGET_LOSS_AMT)
```

```
## [1] 2856.461
```

```
dim(data_amt)
```

```
## [1] 5960   28
```

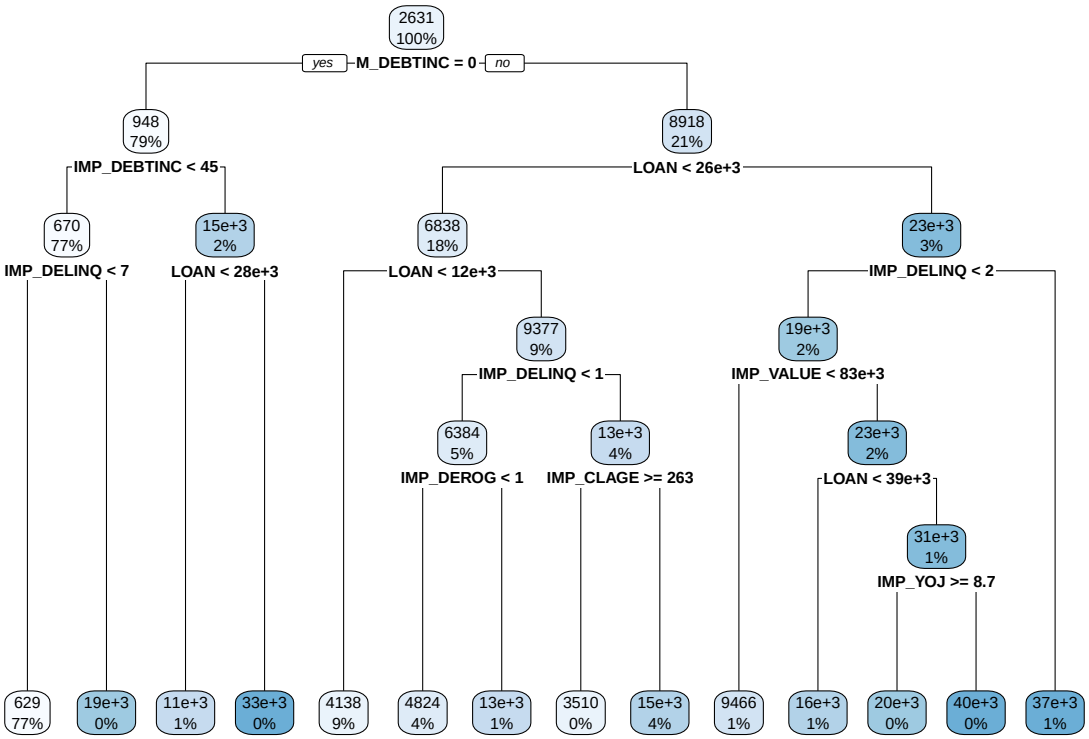
```
dim(data_train_s3)
```

```
## [1] 4772   28
```

```
dim(data_test_s3)
```

```
## [1] 1188   28
```

```
T1A=rpart(data=data_train_s3, TARGET_LOSS_AMT~., control=tr_set, method="anova")
rpart.plot(T1A)
```

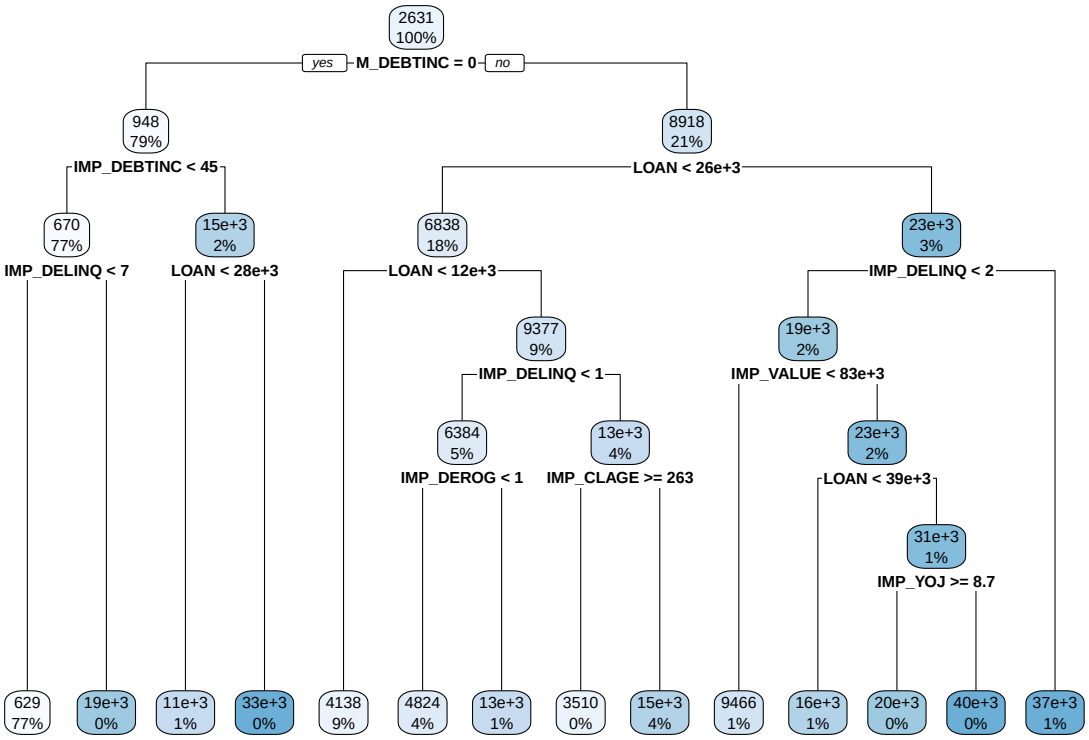


training

```
T1A$variable.importance
```

```
##           M_DEBTINC                LOAN          IMP_DELINQ         IMP_DEBTINC
##          50499939347         47298321377         16230591612         14308366195
##           IMP_VALUE          IMP_MORTDUE           IMP_DEROG           IMP_CLAGE
##          10748121894          8490633455          5037704445          5018137794
##             IMP_CLNO             IMP_YOJ             M_VALUE FLAG.Reason.HomeImp
##           4094152533          3934440475          2229018216          1711445001
## FLAG.Reason.DebtCon             IMP_NINQ             M_DELINQ              M_DEROG
##          1612435786          1371029225          1366064597          1303380183
##              M_NINQ             M_MORTDUE
##           897884126           530864884
```

```
T1P=rpart(data=data_train_s3, TARGET_LOSS_AMT~., control=tr_set, method="anova")
rpart.plot(T1P)
```




```
T1P$variable.importance
```

```
##           M_DEBTINC                LOAN          IMP_DELINQ         IMP_DEBTINC
##         50499939347         47298321377         16230591612         14308366195
##          IMP_VALUE          IMP_MORTDUE            IMP_DEROG           IMP_CLAGE
##         10748121894          8490633455          5037704445          5018137794
##            IMP_CLNO             IMP_YOJ             M_VALUE FLAG.Reason.HomeImp
##          4094152533          3934440475          2229018216          1711445001
## FLAG.Reason.DebtCon             IMP_NINQ            M_DELINQ             M_DEROG
##          1612435786          1371029225          1366064597          1303380183
##              M_NINQ            M_MORTDUE
##           897884126           530864884
```

```
P1A=predict(T1A,data_test_s3)
RMSE1a=sqrt(mean((data_test_s3$TARGET_LOSS_AMT-P1A)^2))

P1P=predict(T1P,data_test_s3)
RMSE1p=sqrt(mean((data_test_s3$TARGET_LOSS_AMT-P1P)^2))
```

```
print(paste("TEST RMSE ANOVA =", RMSE1a))
```

Test

```
## [1] "TEST RMSE ANOVA = 5073.60999691265"
```

```
print(paste("TEST RMSE POISSON =", RMSE1p))
```

```
## [1] "TEST RMSE POISSON = 5073.60999691265"
```

The decision trees are optimal According to the results of my code, both 'anova' and 'poisson' are best but poisson decision trees are better than anova

STEP-4

```
FLAG=sample(c(TRUE,FALSE), nrow(data_amt), replace=TRUE, prob=c(0.8,0.2))
data_train_s4=data[FLAG, ]
data_test_s4=data[! FLAG, ]
```

```
dim(data)
```

1

```
## [1] 5960   29
```

```
dim(data_train_s4)
```

```
## [1] 4791   29
```

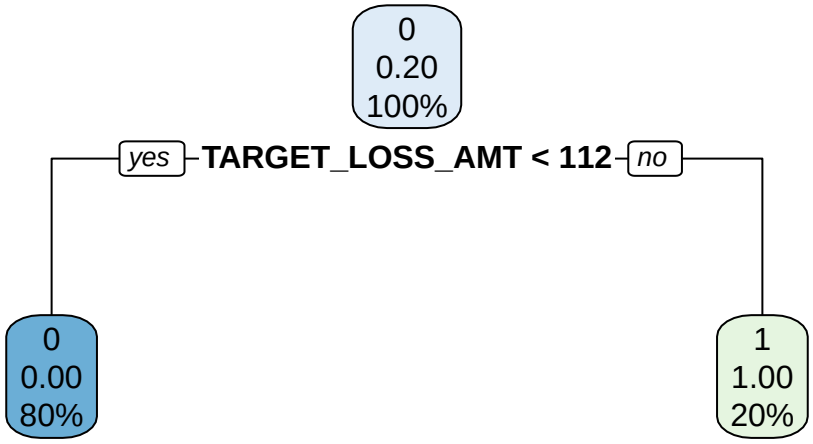
```
dim(data_test_s4)
```

```
## [1] 1169   29
```

```
T1 = rpart(TARGET_BAD_FLAG ~ ., data = data_train_s4, method = "class", parms = list(split = 'gini'))

T2 = rpart(TARGET_LOSS_AMT ~ ., data = subset(data_train_s4, TARGET_BAD_FLAG == 1), method = "anova")
```

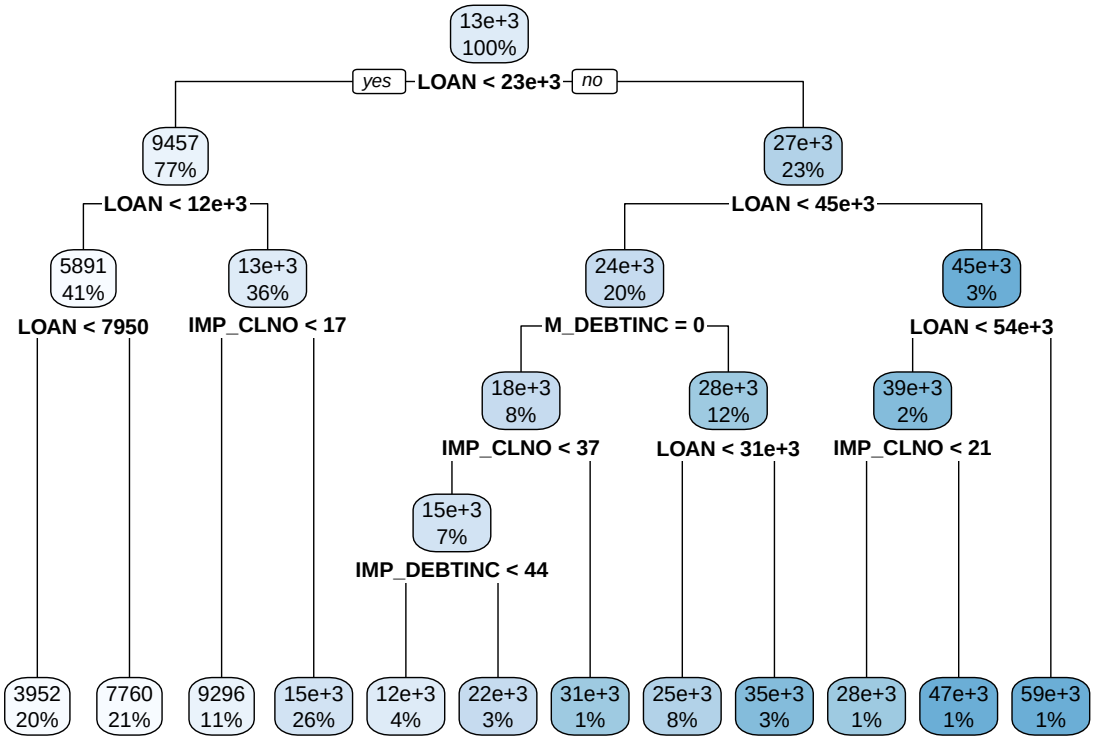
```
rpart.plot(T1)
```




```
T1$variable.importance
```

```
## TARGET_LOSS_AMT       M_DEBTINC      IMP_DELINQ         M_VALUE     IMP_DEBTINC
##      1524.45836       362.27927       118.62242       113.81340       110.60739
##       IMP_DEROG
##        81.75329
```

```
rpart.plot(T2)
```




```
T2$variable.importance
```

```
##               LOAN          IMP_VALUE         IMP_MORTDUE            IMP_CLNO
##         82966851567        13087581144        10026253720         8706925343
##         IMP_DEBTINC          M_DEBTINC FLAG.Reason.HomeImp FLAG.Reason.DebtCon
##          6617818895         4399614784         3267988295         3132098643
##           IMP_CLAGE          IMP_DELINQ           IMP_NINQ       FLAG.Job.Self
##          3004983393         1342714460         1274450249         1179638645
##              IMP_YOJ           IMP_DEROG      FLAG.Job.Sales
##           787924368          295407299          197798597
```

```
pred_prob_default <- predict(T1, data_test_s4, type = "class")
pred_loss_given_default <- predict(T2, data_test_s4)
head(pred_loss_given_default)
```

```
##        3        7        9       15       21       22
## 3951.874 3951.874 3951.874 3951.874 3951.874 3951.874
```

```
pred_prob_default <- predict(T1, data_test_s4, type = "class")
pred_loss_given_default <- predict(T2, data_test_s4)

if (is.factor(pred_loss_given_default)) {
  pred_loss_given_default <- as.numeric(levels(pred_loss_given_default))[pred_loss_given_default]
}
```

```r
pred_loss_given_default[is.na(pred_loss_given_default)] <- 0

pred_prob_default <- as.numeric(as.character(pred_prob_default))
pred_loss_given_default <- as.numeric(pred_loss_given_default)

pred_severity <- pred_prob_default * pred_loss_given_default

head(pred_severity)
```

```
## [1] 3951.874 3951.874 3951.874 3951.874 3951.874 3951.874
```

```r
rmse <- sqrt(mean((data_test$TARGET_LOSS_AMT - pred_severity)^2))
print(rmse)
```

```
## [1] NaN
```

```r
FLAG=sample(c(TRUE,FALSE), nrow(data_amt), replace=TRUE, prob=c(0.7,0.3))
data_train_s4=data[FLAG, ]
data_test_s4=data[! FLAG, ]
```

```r
dim(data)
```

2

```
## [1] 5960   29
```

```r
dim(data_train_s4)
```

```
## [1] 4169   29
```

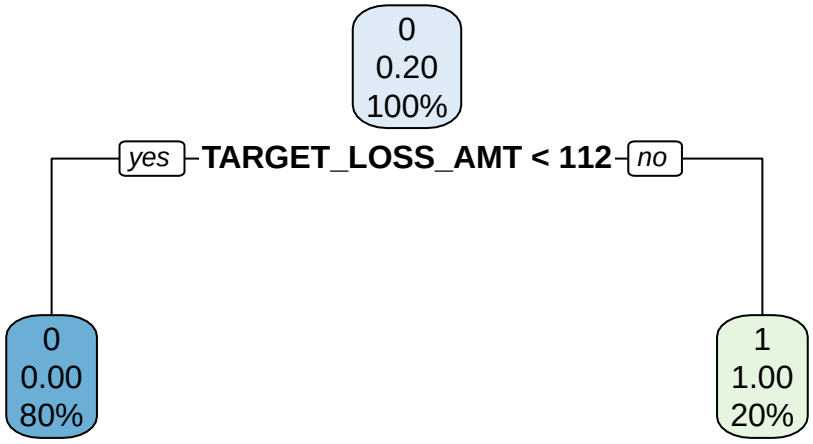
```r
dim(data_test_s4)
```

```
## [1] 1791   29
```

```r
T1 = rpart(TARGET_BAD_FLAG ~ ., data = data_train_s4, method = "class", parms = list(split = 'gini'))

T2 = rpart(TARGET_LOSS_AMT ~ ., data = subset(data_train_s4, TARGET_BAD_FLAG == 1), method = "anova")
```

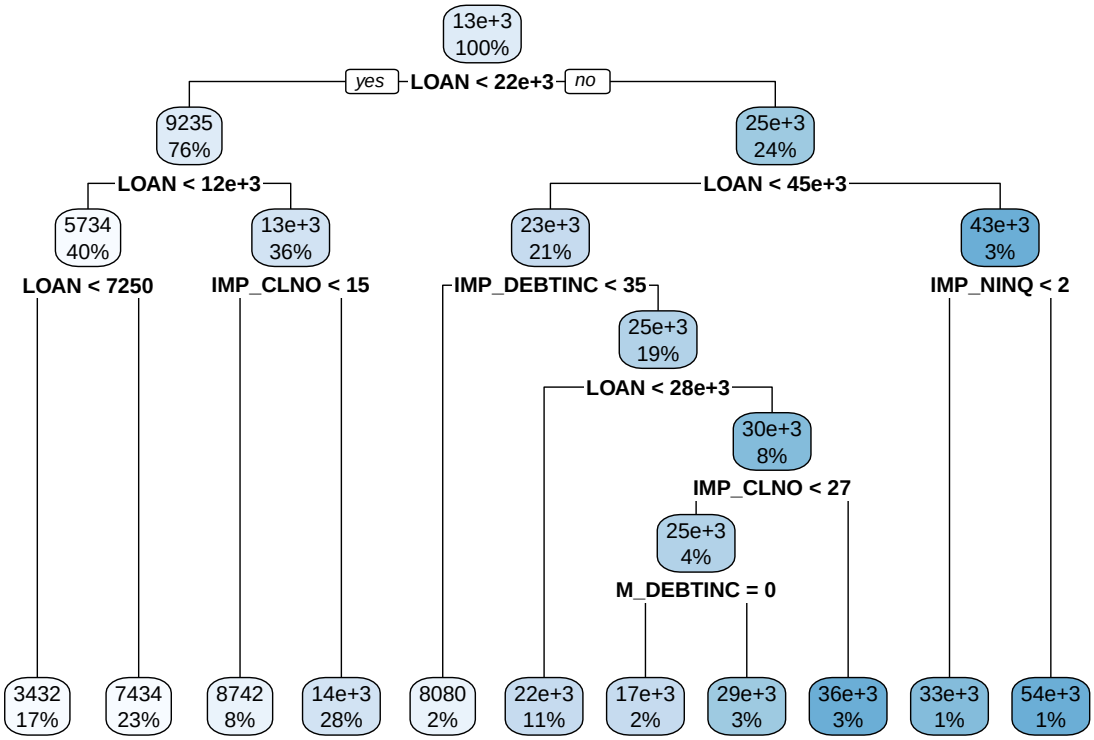
```r
rpart.plot(T1)
```




```
T1$variable.importance
```

```
## TARGET_LOSS_AMT       M_DEBTINC      IMP_DELINQ         M_VALUE     IMP_DEBTINC
##      1346.27009       322.21156       113.25258       111.65747        84.54066
##       IMP_DEROG
##        70.18470
```

```
rpart.plot(T2)
```




```
T2$variable.importance
```

```
##              LOAN         IMP_VALUE       IMP_MORTDUE       IMP_DEBTINC
##       63195425775       10302884999        8971247262        8569104515
##          IMP_CLNO          IMP_NINQ FLAG.Reason.HomeImp FLAG.Reason.DebtCon
##        6517315394        3087265177        2611587559        2418021807
##           IMP_YOJ     FLAG.Job.Other         M_DEBTINC          IMP_CLAGE
##        1392401369        1309071852        1225442880         904803135
##         IMP_DELINQ          IMP_DEROG          M_MORTDUE
##          826591350         226987538          82919351
```

```
pred_prob_default <- predict(T1, data_test_s4, type = "class")
pred_loss_given_default <- predict(T2, data_test_s4)
head(pred_loss_given_default)
```

```
##        4        8        9       13       15       25
## 3432.278 3432.278 3432.278 3432.278 3432.278 3432.278
```

```
pred_prob_default <- predict(T1, data_test_s4, type = "class")
pred_loss_given_default <- predict(T2, data_test_s4)

if (is.factor(pred_loss_given_default)) {
  pred_loss_given_default <- as.numeric(levels(pred_loss_given_default))[pred_loss_given_default]
}
```

```r
pred_loss_given_default[is.na(pred_loss_given_default)] <- 0

pred_prob_default <- as.numeric(as.character(pred_prob_default))
pred_loss_given_default <- as.numeric(pred_loss_given_default)

pred_severity <- pred_prob_default * pred_loss_given_default

head(pred_severity)
```

```
## [1] 3432.278 3432.278 3432.278 3432.278 3432.278 3432.278
```

```r
rmse <- sqrt(mean((data_test$TARGET_LOSS_AMT - pred_severity)^2))
print(rmse)
```

```
## [1] NaN
```

```r
FLAG=sample(c(TRUE,FALSE), nrow(data_amt), replace=TRUE, prob=c(0.8,0.2))
data_train_s4=data[FLAG, ]
data_test_s4=data[! FLAG, ]
```

```r
dim(data)
```

3

```
## [1] 5960   29
```

```r
dim(data_train_s4)
```

```
## [1] 4722   29
```

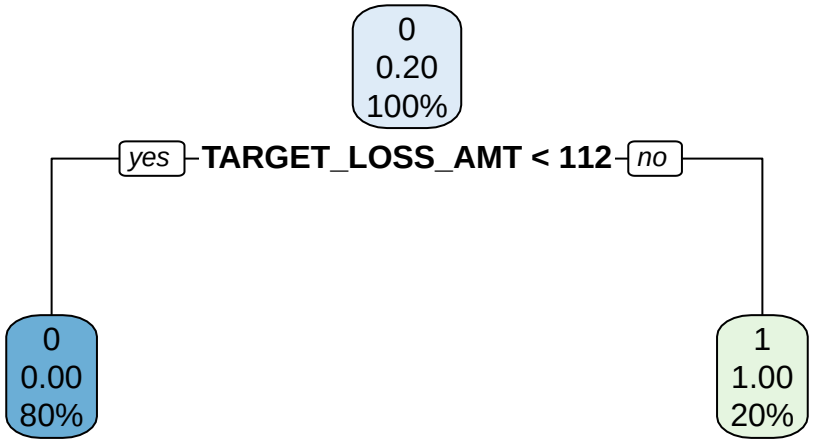
```r
dim(data_test_s4)
```

```
## [1] 1238   29
```

```r
T1 = rpart(TARGET_BAD_FLAG ~ ., data = data_train_s4, method = "class", parms = list(split = 'gini'))

T2 = rpart(TARGET_LOSS_AMT ~ ., data = subset(data_train_s4, TARGET_BAD_FLAG == 1), method = "anova")
```

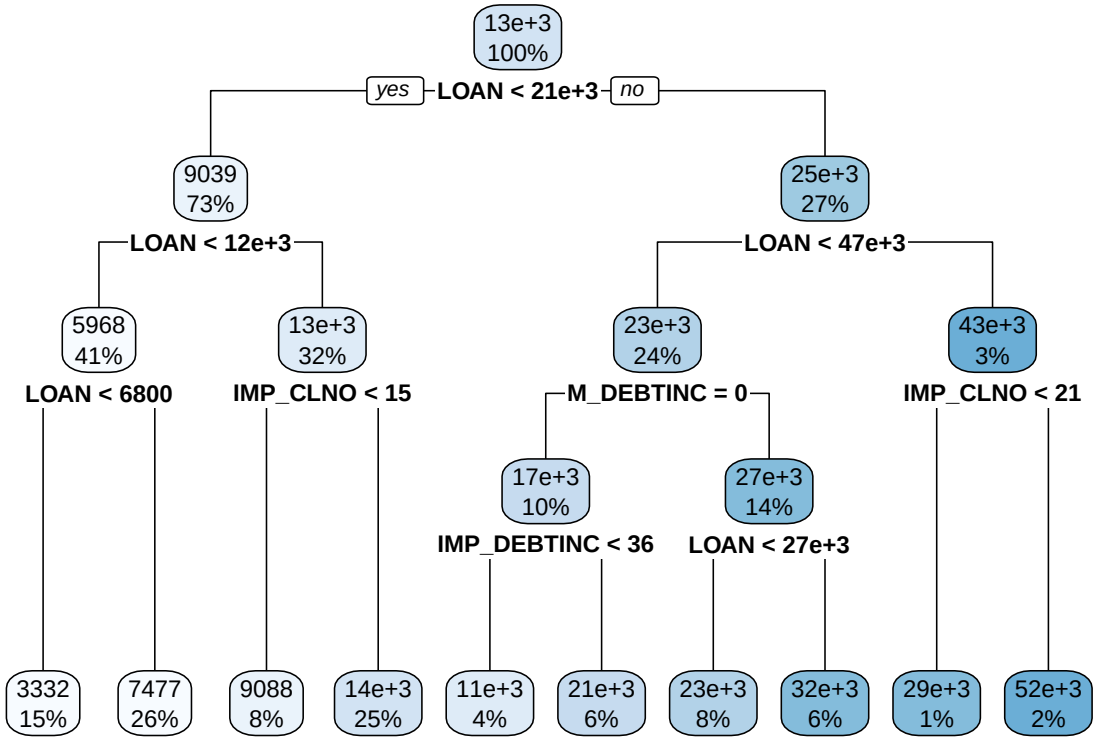
```r
rpart.plot(T1)
```




```
T1$variable.importance
```

```
## TARGET_LOSS_AMT       M_DEBTINC         M_VALUE      IMP_DELINQ     IMP_DEBTINC
##      1510.55993       390.44134       121.61288       120.01271       100.81067
##       IMP_DEROG
##        86.40915
```

```
rpart.plot(T2)
```




```
T2$variable.importance
```

```
##               LOAN          IMP_VALUE         IMP_MORTDUE         IMP_DEBTINC
##         67209722849          9017469127          7921209880          7103383610
##            IMP_CLNO          M_DEBTINC           IMP_CLAGE         FLAG.Job.Mgr
##          6272543620          5433443096          1868215956          1624373155
##           IMP_DELINQ FLAG.Reason.HomeImp FLAG.Reason.DebtCon             M_CLAGE
##          1542825905          1520531936          1454065268          1134590781
##              IMP_YOJ           IMP_NINQ           IMP_DEROG           M_MORTDUE
##           903624993           654297689           209076812            21852781
```

```
pred_prob_default <- predict(T1, data_test_s4, type = "class")
pred_loss_given_default <- predict(T2, data_test_s4)
head(pred_loss_given_default)
```

```
##        3        7       12       25       26       28
## 3332.028 3332.028 3332.028 3332.028 3332.028 3332.028
```

```
pred_prob_default <- predict(T1, data_test_s4, type = "class")
pred_loss_given_default <- predict(T2, data_test_s4)

if (is.factor(pred_loss_given_default)) {
  pred_loss_given_default <- as.numeric(levels(pred_loss_given_default))[pred_loss_given_default]
}
```

```r
pred_loss_given_default[is.na(pred_loss_given_default)] <- 0

pred_prob_default <- as.numeric(as.character(pred_prob_default))
pred_loss_given_default <- as.numeric(pred_loss_given_default)

pred_severity <- pred_prob_default * pred_loss_given_default

head(pred_severity)
```

```
## [1] 3332.028 3332.028 3332.028 3332.028 3332.028 3332.028
```

```r
rmse <- sqrt(mean((data_test$TARGET_LOSS_AMT - pred_severity)^2))
print(rmse)
```

```
## [1] NaN
```

The decision trees are optimal because it is easy to understand. I recommend poisson model because it's decision trees are very easy to understand and it's classifies everything clearly.