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

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

```r
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
```

```r
column_to_exclude_loss <- "TARGET_LOSS_AMT"

updated_dataset_loss <- subset(data, select = -which(names(data) == column_to_exclude_loss))
head(updated_dataset_loss)
```

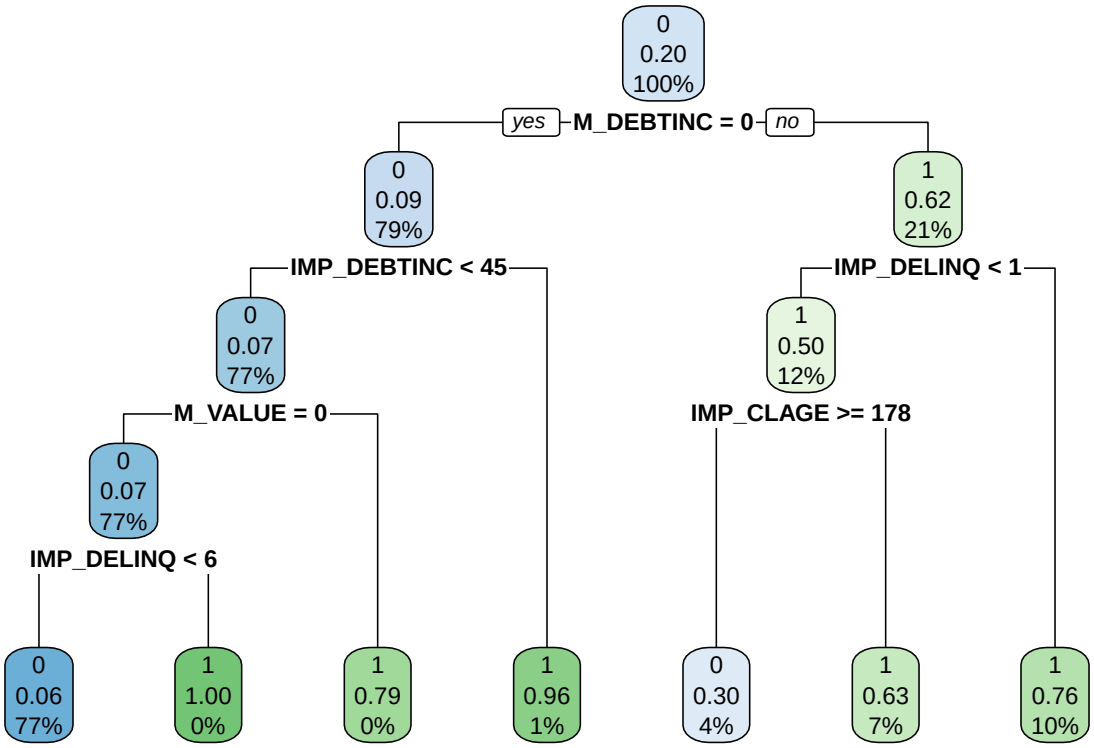
```
##   TARGET_BAD_FLAG LOAN IMP_MORTDUE M_MORTDUE IMP_VALUE M_VALUE IMP_YOJ M_YOJ
## 1               1 1100       25860         0     39025       0    10.5     0
## 2               1 1300       70053         0     68400       0     7.0     0
## 3               1 1500       13500         0     16700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

```
tr_set=rpart.control(maxdepth=10)

t1a_2=rpart(data=updated_dataset_loss, TARGET_BAD_FLAG ~ ., control=tr_set,method="class", parms=list(sp
```

```
rpart.plot(t1a_2)
```



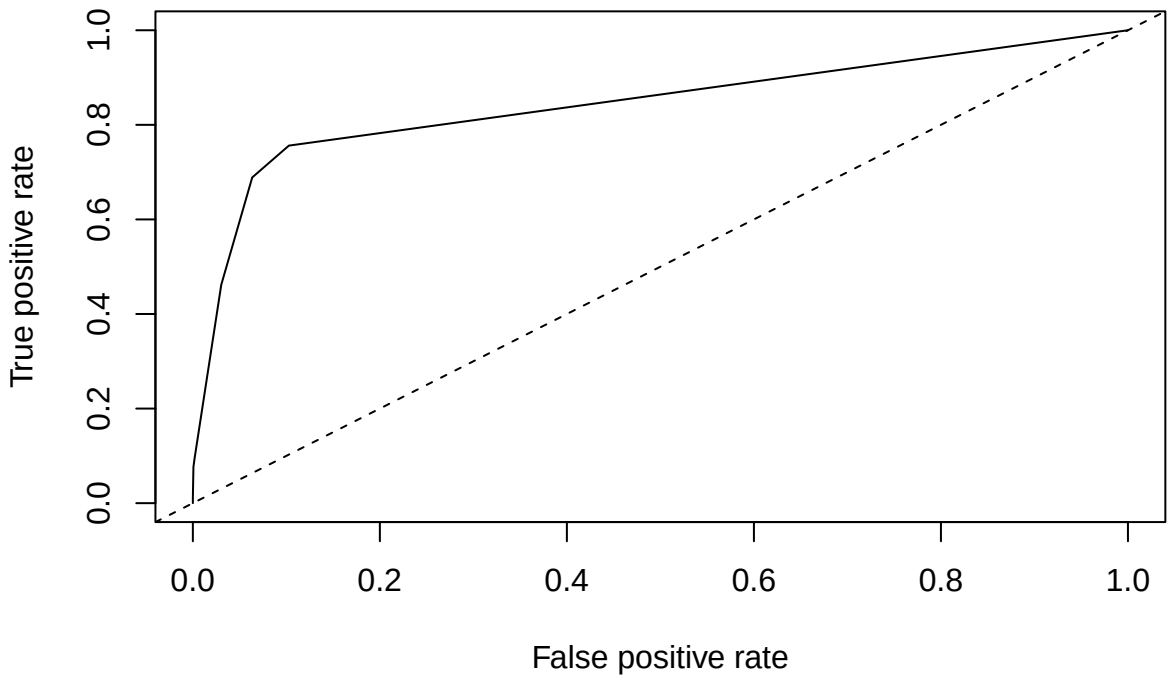

```
as.data.frame(t1a_2$variable.importance)
```

```
##              t1a_2$variable.importance
```

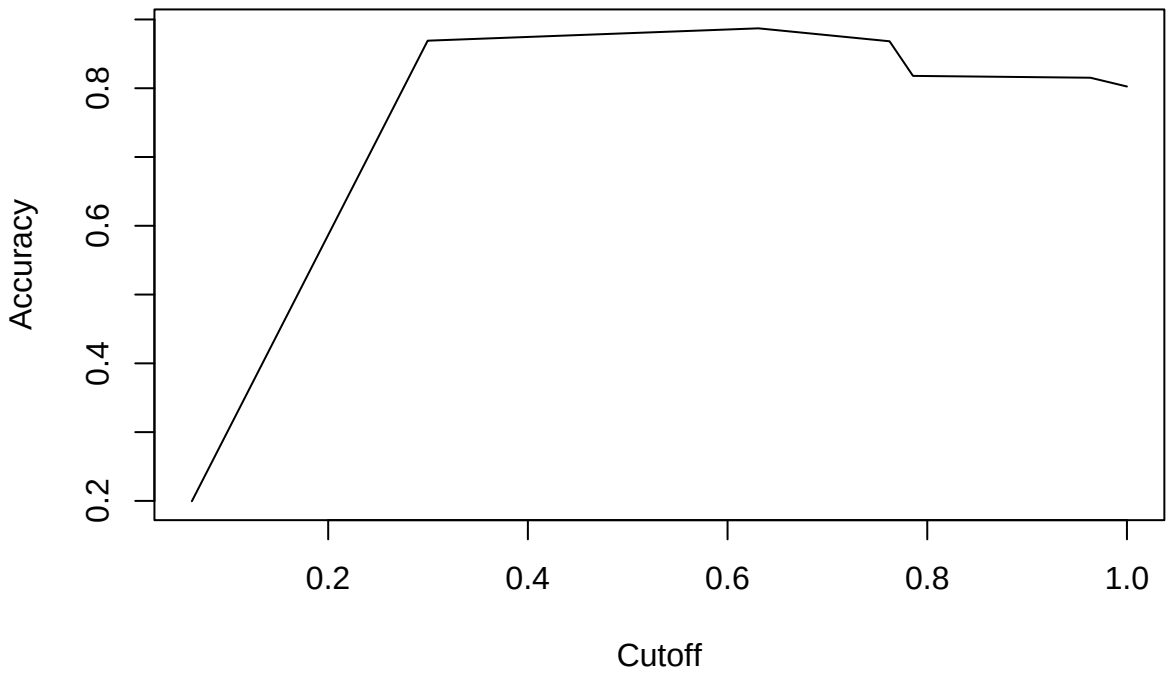
```
## M_DEBTINC                570.021010
## IMP_DEBTINC              128.539072
## IMP_DELINQ                77.371518
## M_VALUE                   51.334486
## IMP_CLAGE                 36.076295
## LOAN                      25.645675
## IMP_DEROG                 22.501563
## M_DEROG                    9.540586
## IMP_VALUE                  8.551021
## M_DELINQ                   7.632469
## M_NINQ                     6.311465
## IMP_YOJ                    4.323751
## M_CLNO                     4.256569
## IMP_CLNO                   2.837461
## IMP_MORTDUE                1.621407
```

```
pred_a2 <- prediction(predict(t1a_2, type = "prob")[, 2], data$TARGET_BAD_FLAG)
```

```
plot(performance(pred_a2, "tpr", "fpr"))
abline(0, 1, lty = 2)
```



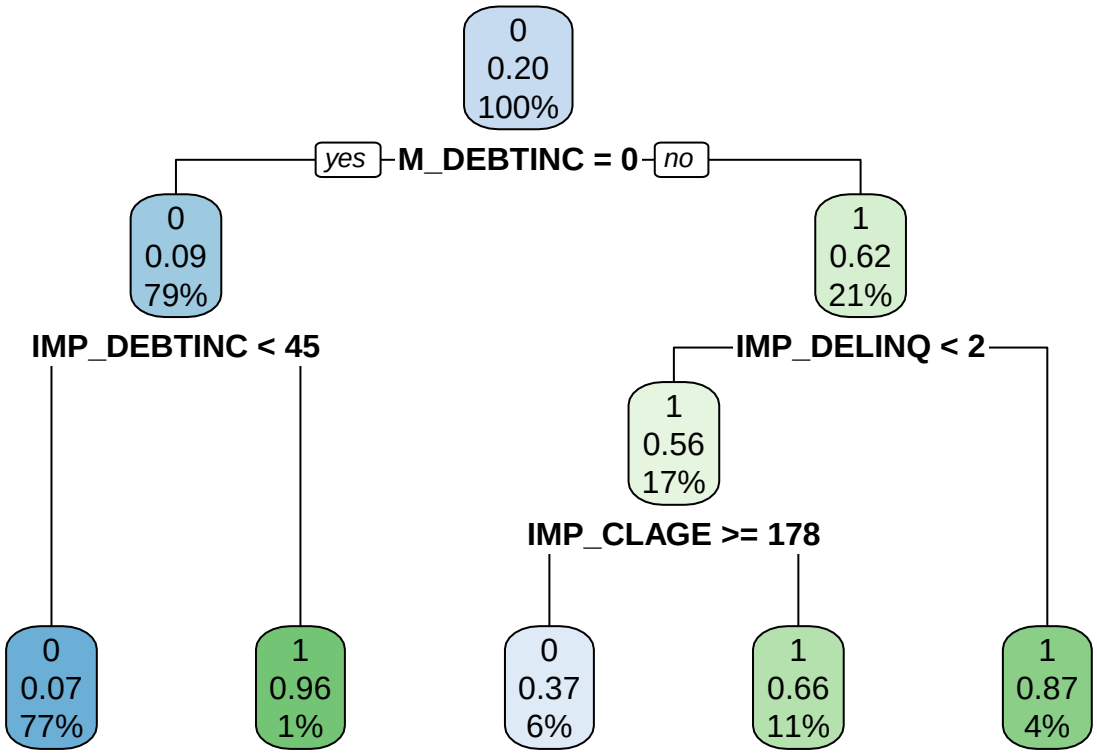

```
plot(performance(pred_a2, "acc"))
```




```
tr_set=rpart.control(maxdepth=10)

t1e_2=rpart(data=updated_dataset_loss, TARGET_BAD_FLAG ~ ., control=tr_set,method="class", parms=list(sp

rpart.plot(t1e_2)
```

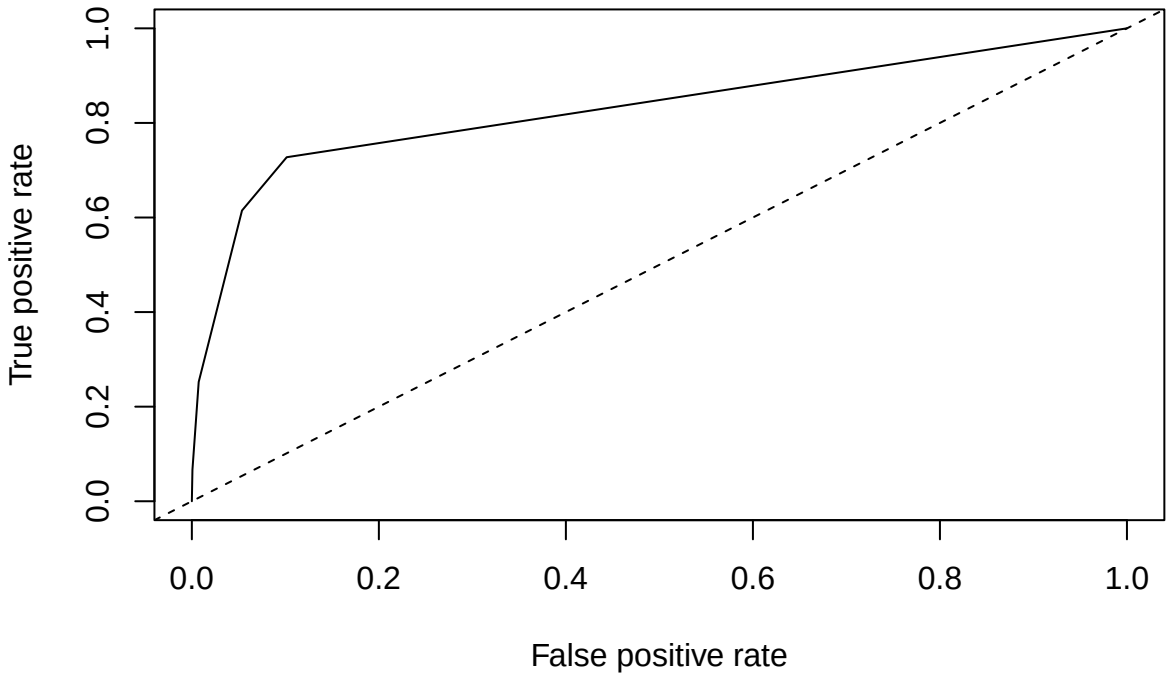



```
as.data.frame(t1e_2$variable.importance)
```

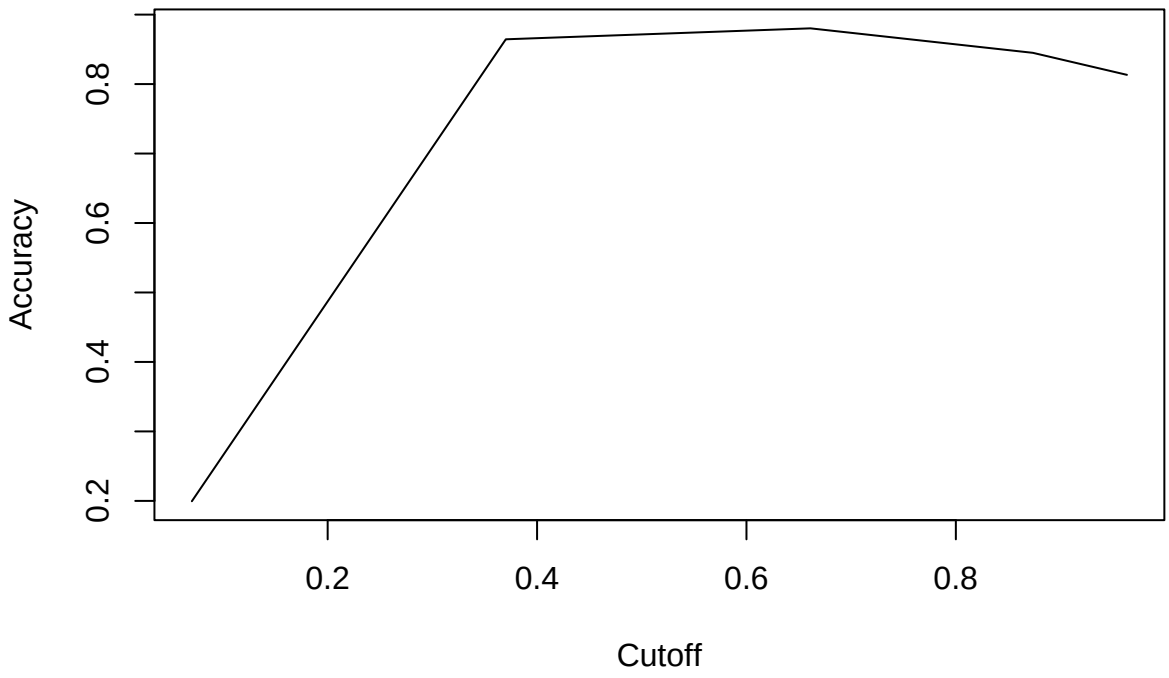
```
##             t1e_2$variable.importance
## M_DEBTINC                   762.591210
## IMP_DEBTINC                 188.922871
## IMP_DELINQ                   68.152477
## IMP_CLAGE                    40.125205
## LOAN                         34.053718
## M_VALUE                      30.094365
## IMP_DEROG                    12.037746
## IMP_VALUE                    10.263083
## IMP_YOJ                       3.436136
## IMP_CLNO                      3.075170
## IMP_MORTDUE                   1.219274
```

```
library("ROCR")
pred_e2 <- prediction(predict(t1e_2, type = "prob")[, 2], data$TARGET_BAD_FLAG)
```

```
plot(performance(pred_e2, "tpr", "fpr"))
abline(0, 1, lty = 2)
```




```
plot(performance(pred_e2, "acc"))
```



#STep-3

```
column_to_exclude_bad <- "TARGET_BAD_FLAG"

updated_dataset_bad <- subset(data, select = -which(names(data) == column_to_exclude_bad))
head(updated_dataset_bad)
```

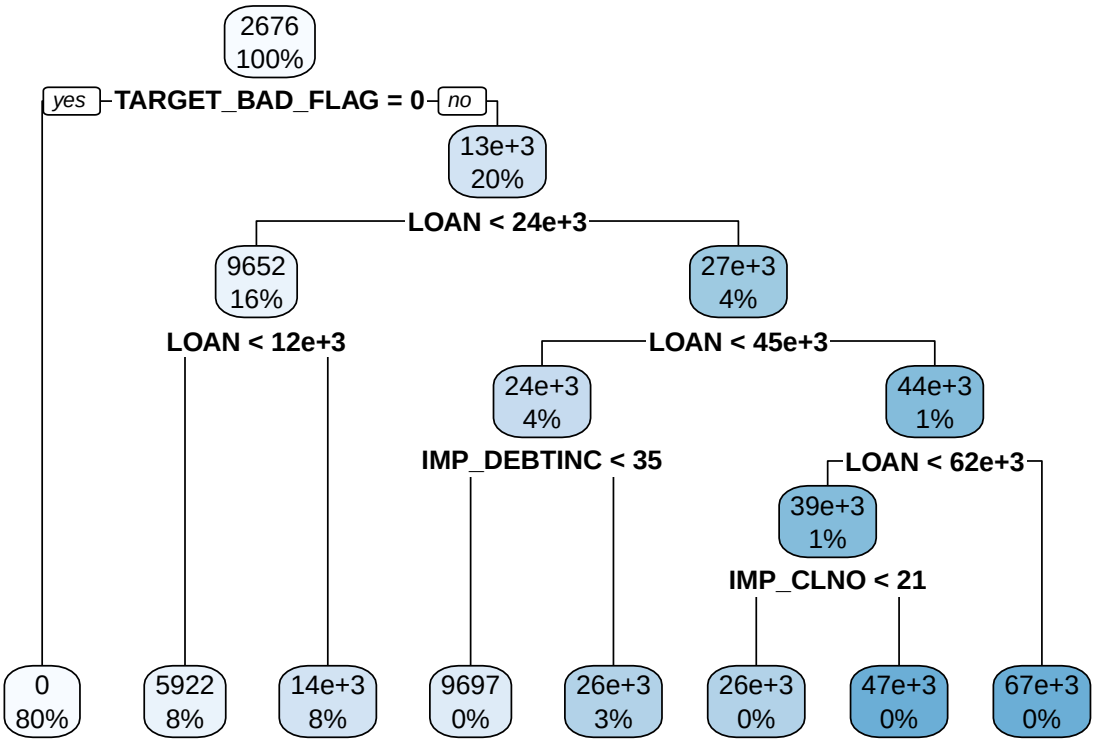
```
##   TARGET_LOSS_AMT LOAN IMP_MORTDUE M_MORTDUE IMP_VALUE M_VALUE IMP_YOJ M_YOJ
## 1             641 1100       25860         0     39025       0    10.5     0
## 2            1109 1300       70053         0     68400       0     7.0     0
## 3             767 1500       13500         0     16700       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
```

```r
tr_set=rpart.control(maxdepth=10)

t1a_3=rpart(data=data, TARGET_LOSS_AMT ~ ., control=tr_set,method="anova")
```

```r
rpart.plot(t1a_3)
```



```r
as.data.frame(t1a_3$variable.importance)
```

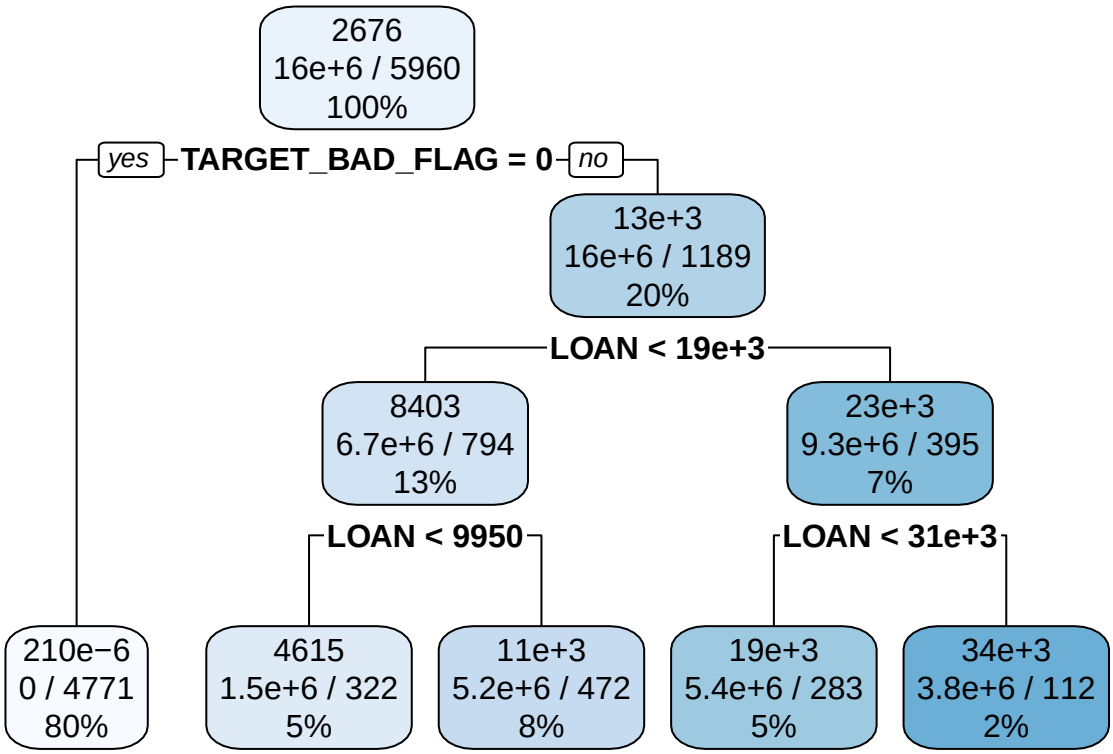
```
##                      t1a_3$variable.importance
## TARGET_BAD_FLAG                  171276970985
## LOAN                              93651039516
## M_DEBTINC                         43935640160
## IMP_DEBTINC                       19117549170
## IMP_DELINQ                        17005021630
## M_VALUE                           14117025363
## IMP_VALUE                         13002823330
## IMP_MORTDUE                       10119356810
## IMP_DEROG                          9795486987
## IMP_CLNO                           5696550973
## FLAG.Reason.HomeImp                4166142650
## FLAG.Reason.DebtCon                3945210843
## IMP_NINQ                           2339775381
## IMP_CLAGE                          1374109492
## IMP_YOJ                             731959747
## M_MORTDUE                           620243894
```

```r
p1a_3=predict(t1a_3,data)
```

```r
RMSE1a_3= sqrt(mean((data$TARGET_LOSS_AMT- p1a_3)^2))
```

```r
t1p_3=rpart(data=data, TARGET_LOSS_AMT ~ ., control=tr_set,method="poisson")
```

```r
rpart.plot(t1p_3
           )
```




```
as.data.frame(t1p_3$variable.importance)
```

```
##                       t1p_3$variable.importance
## TARGET_BAD_FLAG                    51421236.62
## M_DEBTINC                          13190477.01
## LOAN                                5868622.63
## IMP_DELINQ                          4528081.84
## M_VALUE                             4238251.63
## IMP_DEBTINC                         3402453.35
## IMP_DEROG                           2947221.47
## IMP_VALUE                           1373149.60
## IMP_MORTDUE                         1014807.22
## FLAG.Reason.HomeImp                  204922.84
## FLAG.Reason.DebtCon                  164559.25
## FLAG.Job.Self                        147185.77
## IMP_YOJ                               31969.03
## IMP_CLNO                              24839.13
## IMP_CLAGE                             19181.42
```

```
p1p_3=predict(t1p_3,data)
```

```
RMSE1p_3= sqrt(mean((data$TARGET_LOSS_AMT- p1p_3)^2))
print(RMSE1p_3)
```

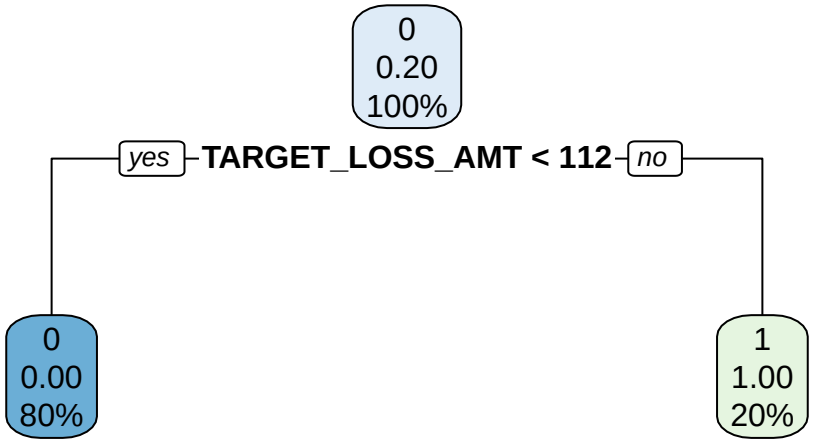
```
## [1] 3001.205
```

######Step-4

```
tr_set=rpart.control(maxdepth=10)

t1a_4=rpart(data=data, TARGET_BAD_FLAG ~ ., control=tr_set,method="class", parms=list(split="gini"))
```

```
rpart.plot(t1a_4)
```



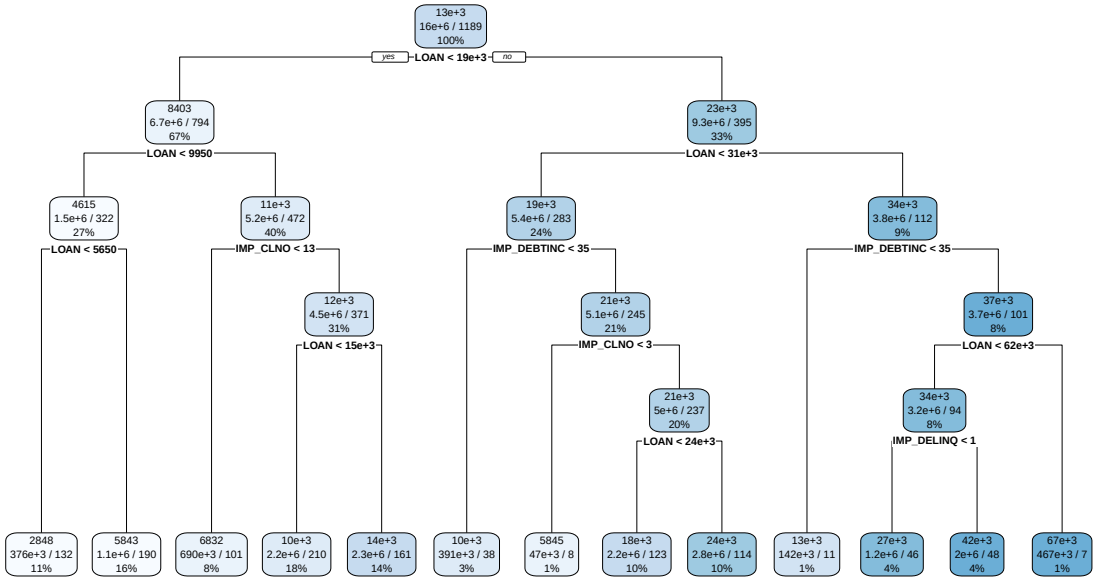

```
pred_a4 <- prediction(predict(t1a_4, type = "prob")[, 2], data$TARGET_BAD_FLAG)
as.data.frame(t1a_4$variable.importance)
```

```
##                 t1a_4$variable.importance
## TARGET_LOSS_AMT                 1903.5970
## M_DEBTINC                        488.3070
## IMP_DELINQ                       164.9037
## M_VALUE                          156.8987
## IMP_DEBTINC                      121.6765
## IMP_DEROG                        108.8685
```

```
bad_flag_1_data <- subset(data, TARGET_BAD_FLAG == 1)
```

```
t1p_4=rpart(data=bad_flag_1_data, TARGET_LOSS_AMT ~ ., control=tr_set,method="poisson")
```

```
rpart.plot(t1p_4)
```




```
as.data.frame(t1p_4$variable.importance)
```

```
##                       t1p_4$variable.importance
## LOAN                                6409665.00
## IMP_VALUE                           1481448.38
## IMP_MORTDUE                         1081934.14
## IMP_DEBTINC                          574282.90
## IMP_CLNO                             446748.28
## FLAG.Reason.HomeImp                  229285.80
## IMP_DELINQ                           223669.68
## FLAG.Reason.DebtCon                  188922.21
## FLAG.Job.Self                        147185.77
## IMP_CLAGE                             51185.99
## IMP_NINQ                              48213.49
## IMP_DEROG                             45544.27
## IMP_YOJ                               38733.92
## M_VALUE                               12118.28
## FLAG.Job.Other                         7457.40
```

```
probability_of_default <- predict(t1a_4, type = "prob")
```

```r
p1p_4=predict(t1p_4,data)
```

```r
probability_severity_model <- probability_of_default * p1p_4
```

```r
rmse_probability_severity <- sqrt(mean((probability_severity_model - data$TARGET_LOSS_AMT)^2))
print(rmse_probability_severity)
```

```
## [1] 11188.29
```