Random-forest

Data Set Information: 2126 fetal cardiotocograms (CTGs) were automatically processed and the respective diagnostic features measured. The CTGs were also classified by three expert obstetricians and a consensus classification label assigned to each of them. NSP: 1. Normal, 2. Suspected, 3. pathology

```
data <-
read.csv("file:///C:/Users/badal/Desktop/datset_/Cardiotocographic.csv")
head(data)
```

```
##    LB          AC FM          UC          DL DS          DP ASTV MSTV ALTV
MLTV
## 1 120 0.000000000  0 0.000000000 0.000000000  0 0.000000000   73  0.5   43
2.4
## 2 132 0.006379585  0 0.006379585 0.003189793  0 0.000000000   17  2.1    0
10.4
## 3 133 0.003322259  0 0.008305648 0.003322259  0 0.000000000   16  2.1    0
13.4
## 4 134 0.002560819  0 0.007682458 0.002560819  0 0.000000000   16  2.4    0
23.0
## 5 132 0.006514658  0 0.008143322 0.000000000  0 0.000000000   16  2.4    0
19.9
## 6 134 0.001049318  0 0.010493179 0.009443861  0 0.002098636   26  5.9    0
0.0
##   Width Min Max Nmax Nzeros Mode Mean Median Variance Tendency NSP
## 1    64  62 126    2      0  120  137    121       73        1   2
## 2   130  68 198    6      1  141  136    140       12        0   1
## 3   130  68 198    5      1  141  135    138       13        0   1
## 4   117  53 170   11      0  137  134    137       13        1   1
## 5   117  53 170    9      0  137  136    138       11        1   1
## 6   150  50 200    5      3   76  107    107      170        0   3
```

```
str(data)
```

```
## 'data.frame':    2126 obs. of  22 variables:
##  $ LB      : int  120 132 133 134 132 134 134 122 122 122 ...
##  $ AC      : num  0 0.00638 0.00332 0.00256 0.00651 ...
##  $ FM      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ UC      : num  0 0.00638 0.00831 0.00768 0.00814 ...
##  $ DL      : num  0 0.00319 0.00332 0.00256 0 ...
##  $ DS      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ DP      : num  0 0 0 0 0 ...
##  $ ASTV    : int  73 17 16 16 16 26 29 83 84 86 ...
##  $ MSTV    : num  0.5 2.1 2.1 2.4 2.4 5.9 6.3 0.5 0.5 0.3 ...
##  $ ALTV    : int  43 0 0 0 0 0 0 6 5 6 ...
##  $ MLTV    : num  2.4 10.4 13.4 23 19.9 0 0 15.6 13.6 10.6 ...
##  $ Width   : int  64 130 130 117 117 150 150 68 68 68 ...
```

```
##  $ Min     : int  62 68 68 53 53 50 50 62 62 62 ...
##  $ Max     : int  126 198 198 170 170 200 200 130 130 130 ...
##  $ Nmax    : int  2 6 5 11 9 5 6 0 0 1 ...
##  $ Nzeros  : int  0 1 1 0 0 3 3 0 0 0 ...
##  $ Mode    : int  120 141 141 137 137 76 71 122 122 122 ...
##  $ Mean    : int  137 136 135 134 136 107 107 122 122 122 ...
##  $ Median  : int  121 140 138 137 138 107 106 123 123 123 ...
##  $ Variance: int  73 12 13 13 11 170 215 3 3 1 ...
##  $ Tendency: int  1 0 0 1 1 0 0 1 1 1 ...
##  $ NSP     : int  2 1 1 1 1 3 3 3 3 3 ...
```

```
data$NSP <- as.factor(data$NSP)
```

```
summary(data)
```

```
##        LB              AC                  FM                  UC          
##  Min.   :106.0   Min.   :0.000000   Min.   :0.000000   Min.   :0.000000  
##  1st Qu.:126.0   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.001876  
##  Median :133.0   Median :0.001630   Median :0.000000   Median :0.004482  
##  Mean   :133.3   Mean   :0.003170   Mean   :0.009474   Mean   :0.004357  
##  3rd Qu.:140.0   3rd Qu.:0.005631   3rd Qu.:0.002512   3rd Qu.:0.006525  
##  Max.   :160.0   Max.   :0.019284   Max.   :0.480634   Max.   :0.014925  
##        DL                  DS                  DP                 ASTV       
##  Min.   :0.000000   Min.   :0.000e+00   Min.   :0.0000000   Min.   :12.00  
##  1st Qu.:0.000000   1st Qu.:0.000e+00   1st Qu.:0.0000000   1st Qu.:32.00  
##  Median :0.000000   Median :0.000e+00   Median :0.0000000   Median :49.00  
##  Mean   :0.001885   Mean   :3.585e-06   Mean   :0.0001566   Mean   :46.99  
##  3rd Qu.:0.003264   3rd Qu.:0.000e+00   3rd Qu.:0.0000000   3rd Qu.:61.00  
##  Max.   :0.015385   Max.   :1.353e-03   Max.   :0.0053476   Max.   :87.00  
##       MSTV             ALTV              MLTV            Width       
##  Min.   :0.200   Min.   : 0.000   Min.   : 0.000   Min.   :  3.00  
##  1st Qu.:0.700   1st Qu.: 0.000   1st Qu.: 4.600   1st Qu.: 37.00  
##  Median :1.200   Median : 0.000   Median : 7.400   Median : 67.50  
##  Mean   :1.333   Mean   : 9.847   Mean   : 8.188   Mean   : 70.45  
##  3rd Qu.:1.700   3rd Qu.:11.000   3rd Qu.:10.800   3rd Qu.:100.00  
##  Max.   :7.000   Max.   :91.000   Max.   :50.700   Max.   :180.00  
##       Min              Max            Nmax             Nzeros       
##  Min.   : 50.00   Min.   :122   Min.   : 0.000   Min.   : 0.0000  
##  1st Qu.: 67.00   1st Qu.:152   1st Qu.: 2.000   1st Qu.: 0.0000  
##  Median : 93.00   Median :162   Median : 3.000   Median : 0.0000  
##  Mean   : 93.58   Mean   :164   Mean   : 4.068   Mean   : 0.3236  
##  3rd Qu.:120.00   3rd Qu.:174   3rd Qu.: 6.000   3rd Qu.: 0.0000  
##  Max.   :159.00   Max.   :238   Max.   :18.000   Max.   :10.0000  
##       Mode            Mean           Median         Variance      
##  Min.   : 60.0   Min.   : 73.0   Min.   : 77.0   Min.   :  0.00  
##  1st Qu.:129.0   1st Qu.:125.0   1st Qu.:129.0   1st Qu.:  2.00  
##  Median :139.0   Median :136.0   Median :139.0   Median :  7.00  
##  Mean   :137.5   Mean   :134.6   Mean   :138.1   Mean   : 18.81  
##  3rd Qu.:148.0   3rd Qu.:145.0   3rd Qu.:148.0   3rd Qu.: 24.00  
##  Max.   :187.0   Max.   :182.0   Max.   :186.0   Max.   :269.00
```

```
##     Tendency       NSP
##  Min.   :-1.0000   1:1655
##  1st Qu.: 0.0000   2: 295
##  Median : 0.0000   3: 176
##  Mean   : 0.3203
##  3rd Qu.: 1.0000
##  Max.   : 1.0000

table(data$NSP)

##
##    1    2    3
## 1655  295  176
```

partition data into traning and validation sets

```
set.seed(1234)
index <- sample(2, nrow(data), replace = T, prob = c(0.70, 0.30))
train <- data[index==1,]
validate <- data[index==2,]
```

Random forest model:

```
#install.packages("randomForest")
library(randomForest)

## Warning: package 'randomForest' was built under R version 3.6.3

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

set.seed(111)
rf<-randomForest(NSP~., data = train)
rf

##
## Call:
##  randomForest(formula = NSP ~ ., data = train)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 4
##
##         OOB estimate of  error rate: 5.84%
## Confusion matrix:
##      1   2   3 class.error
## 1 1175  17   3  0.01673640
## 2   51 144   6  0.28358209
## 3    6   6 115  0.09448819

summary(rf) #attributes of rf
```

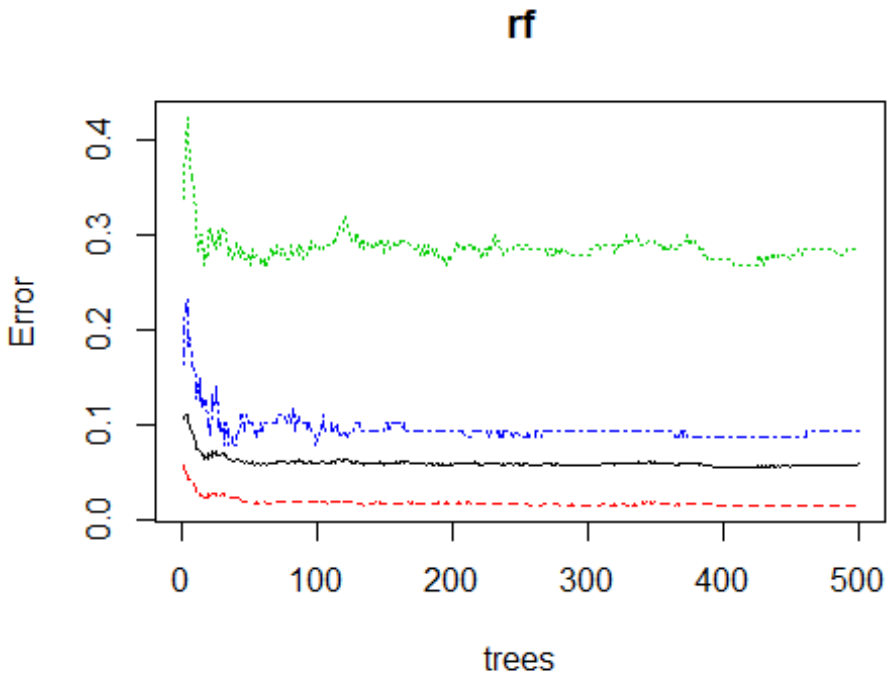
```
##                 Length Class  Mode
## call               3   -none- call
## type               1   -none- character
## predicted       1523   factor numeric
## err.rate        2000   -none- numeric
## confusion         12   -none- numeric
## votes           4569   matrix numeric
## oob.times       1523   -none- numeric
## classes            3   -none- character
## importance        21   -none- numeric
## importanceSD       0   -none- NULL
## localImportance    0   -none- NULL
## proximity          0   -none- NULL
## ntree              1   -none- numeric
## mtry               1   -none- numeric
## forest            14   -none- list
## y               1523   factor numeric
## test               0   -none- NULL
## inbag              0   -none- NULL
## terms              3   terms  call
```

```
rf$confusion
```

```
##      1   2   3 class.error
## 1 1175  17   3  0.01673640
## 2   51 144   6  0.28358209
## 3    6   6 115  0.09448819
```

Error rate

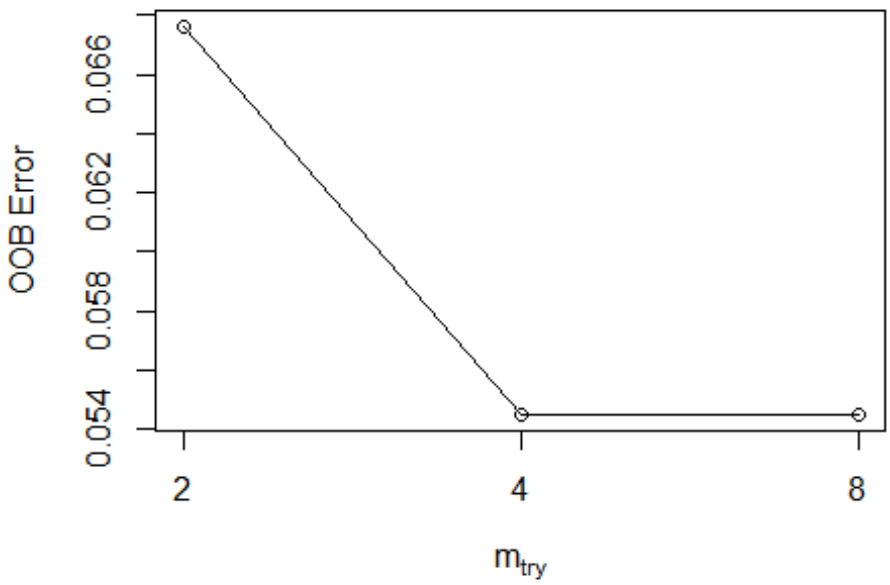
```
plot(rf)
```



Tune random forest model

```
tuneRF(train[,-22], train$NSP,
       stepFactor = 0.5,
       plot = TRUE,
       ntreeTry = 300,
       trace = TRUE,
       improve = 0.05
       )
```

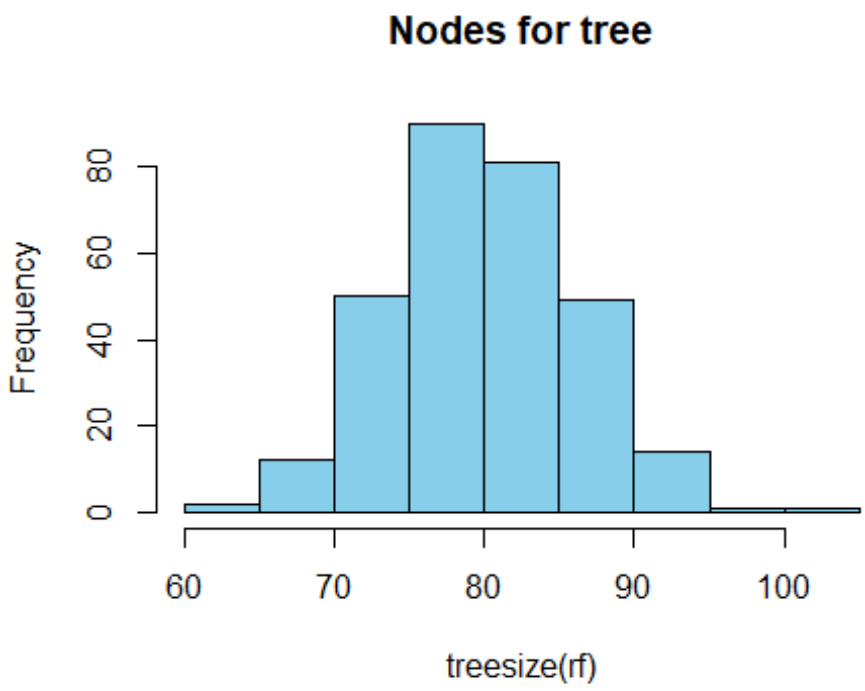
```
## mtry = 4  OOB error = 5.45%
## Searching left ...
## mtry = 8     OOB error = 5.45%
## 0 0.05
## Searching right ...
## mtry = 2     OOB error = 6.76%
## -0.2409639 0.05
```



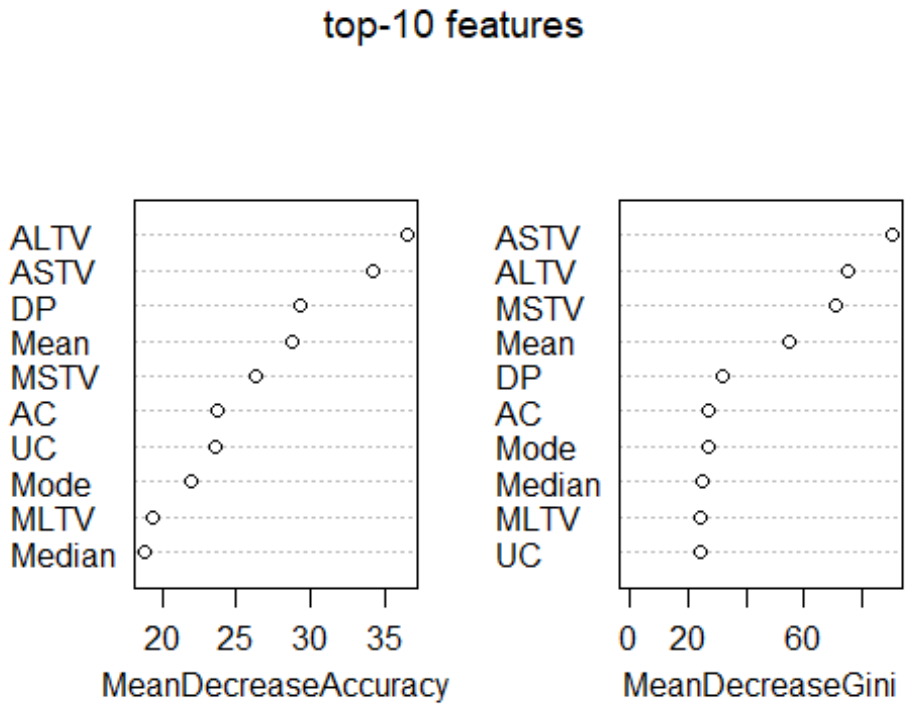

```
##       mtry   OOBError
## 2.OOB    2 0.06762968
## 4.OOB    4 0.05449770
## 8.OOB    8 0.05449770

set.seed(222)
rf<-randomForest(NSP~., data = train,
                 ntree = 300,
                 mtry = 8,
                 importance = TRUE,
                 proximity = TRUE)
rf

##
## Call:
##  randomForest(formula = NSP ~ ., data = train, ntree = 300, mtry = 8,
importance = TRUE, proximity = TRUE)
##                Type of random forest: classification
##                      Number of trees: 300
## No. of variables tried at each split: 8
##
##         OOB estimate of  error rate: 5.58%
## Confusion matrix:
##      1   2   3 class.error
## 1 1172  18   5  0.01924686
## 2   51 147   3  0.26865672
## 3    5   3 119  0.06299213
```

Number of nodes for the tree

```
hist(treesize(rf), main = "Nodes for tree",
     col = "skyblue")
```




```
varImpPlot(rf, sort =TRUE, n.var =10, main = "top-10 features", cex = 1)
```



graph 1 : tests how worse the model performs without each variable. graph 2: tells us how pure the node are at the end of the tree without rach variable.

quantative values.

```
importance(rf) #gives Gini Index (priority of variable)
```

```
##                  1           2         3 MeanDecreaseAccuracy
MeanDecreaseGini
## LB       17.132286  7.16274976  7.915169            18.805473
18.02488392
## AC       21.100620 17.98906880 11.200648            23.698048
27.22863814
## FM        9.744602 10.39940650  2.186584            13.131800
9.73630145
## UC       12.705099 17.20023502 16.931399            23.630715
24.11442412
## DL        3.083279  1.35286211  5.486728             6.098206
2.98148979
## DS        1.001671  0.00000000  1.001671             1.418872
0.04737006
## DP       27.312479  8.34349989 16.750946            29.359711
31.80676618
## ASTV     18.921616 32.72507442 31.466362            34.225644
90.53128264
## MSTV     14.854489 23.26088999 21.324683            26.363290
70.92520421
## ALTV     23.305467 26.53479445 33.902552            36.558538
```

```
74.86578357
## MLTV     12.360262 12.71562783  9.990416           19.410874
24.36258107
## Width    13.333146  4.41577766  5.059051           14.839750
12.88031946
## Min      10.621741  5.63112208  8.668200           15.052170
14.10036320
## Max      11.920657  5.81771565  6.494584           15.013933
13.25917462
## Nmax      8.490765  2.61552171  4.776950            9.589162
8.94732671
## Nzeros    3.498580  1.86340874  2.545108            4.551117
2.00771437
## Mode     17.060443  8.53445890 10.136993           21.992184
26.93020145
## Mean     23.503174 11.56692165 19.502985           28.809405
55.15406327
## Median   15.417586 10.14251823 10.577731           18.873297
24.60305914
## Variance 11.838876  1.90823463  6.554857           12.023446
10.53096562
## Tendency  4.870128 -0.01828876  4.359204            6.385997
2.71653751
```

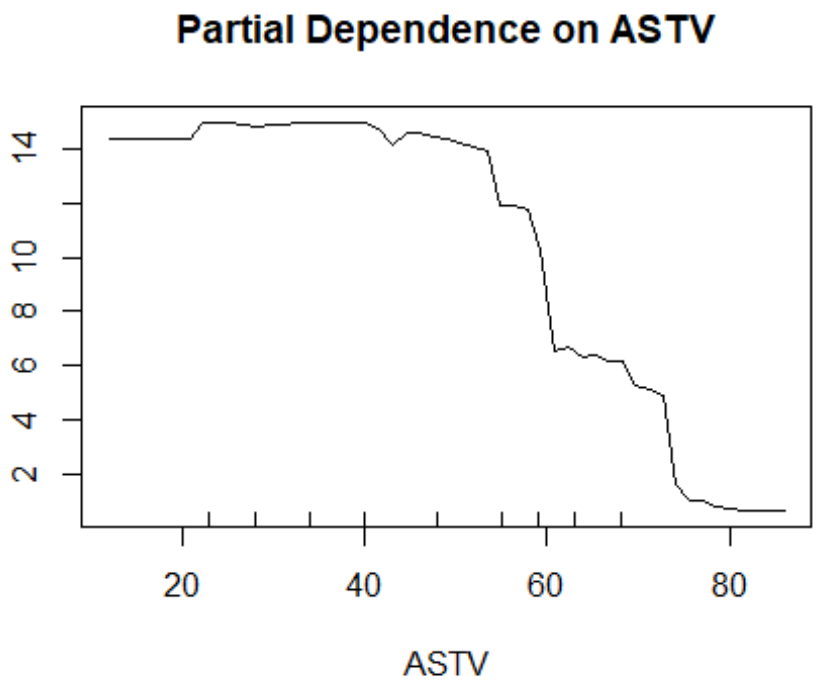
To find out which predictor variables are actully used in the random forest.

```
varUsed(rf)
```

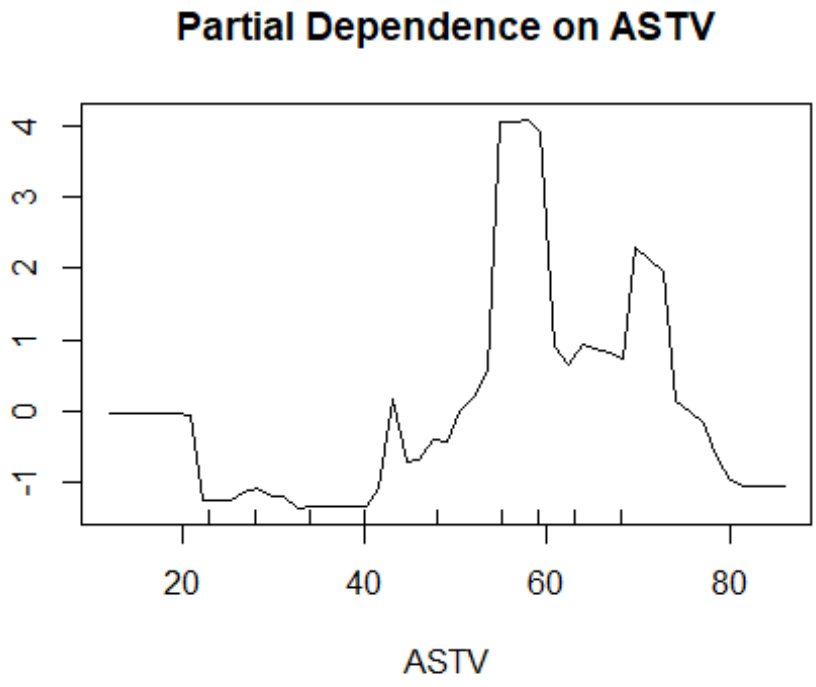
```
##  [1] 1275 1058 1011 1691  331    4  696 2203 1220 2151 1659 1248 1374 1201
856
## [16]  278 1384 1589 1300  953  351
```

Partial dependance plot.

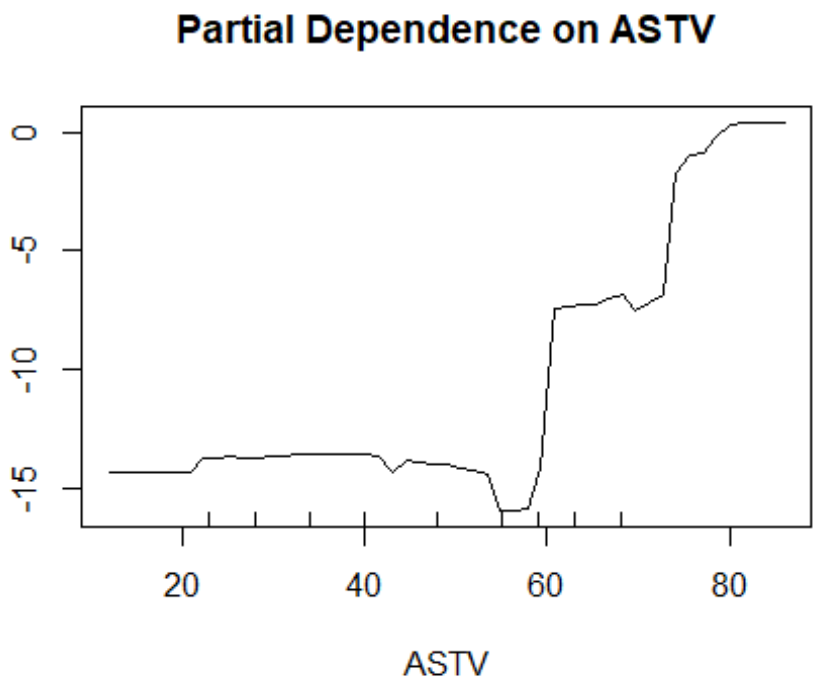
```
partialPlot(rf, train, ASTV, "1")
```




```
partialPlot(rf, train, ASTV, "2")
```




```
partialPlot(rf, train, ASTV, "3")
```



Extract Single Tree from the forest:

```
getTree(rf,1,labelVar = TRUE)
```

```
##     left daughter right daughter split var  split point status prediction
## 1               2              3      ALTV 1.650000e+01      1       <NA>
## 2               4              5        LB 1.485000e+02      1       <NA>
## 3               6              7      ASTV 5.950000e+01      1       <NA>
## 4               8              9        DP 1.563723e-03      1       <NA>
## 5              10             11      Nmax 5.000000e-01      1       <NA>
## 6              12             13      ALTV 5.150000e+01      1       <NA>
## 7              14             15        FM 2.843826e-03      1       <NA>
## 8              16             17      Mode 9.250000e+01      1       <NA>
## 9              18             19      Mean 1.100000e+02      1       <NA>
## 10              0              0      <NA> 0.000000e+00     -1          2
## 11             20             21        AC 2.934742e-03      1       <NA>
## 12             22             23      Mode 1.610000e+02      1       <NA>
## 13             24             25      MLTV 8.300000e+00      1       <NA>
## 14             26             27       Max 1.545000e+02      1       <NA>
## 15             28             29      MLTV 4.050000e+00      1       <NA>
## 16              0              0      <NA> 0.000000e+00     -1          3
## 17             30             31      ASTV 7.750000e+01      1       <NA>
## 18             32             33      Nmax 2.500000e+00      1       <NA>
## 19             34             35        UC 1.858616e-03      1       <NA>
## 20             36             37       Max 1.780000e+02      1       <NA>
## 21              0              0      <NA> 0.000000e+00     -1          1
## 22             38             39      MLTV 1.060000e+01      1       <NA>
```

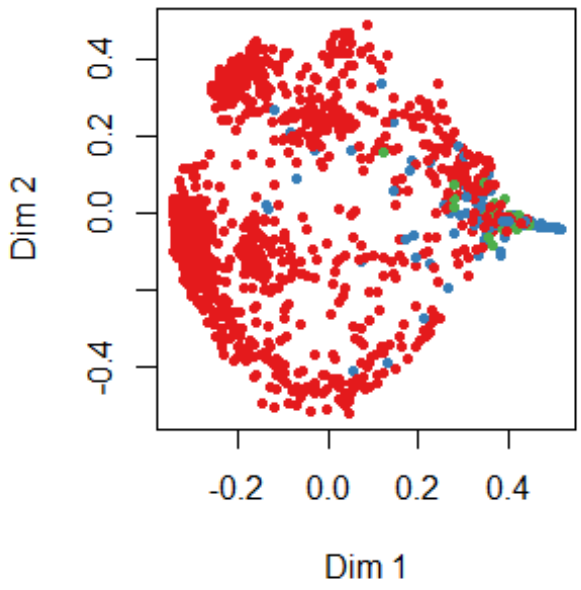
```
## 23              40              41     Width 9.600000e+01      1       <NA>
## 24              42              43    Median 1.595000e+02      1       <NA>
## 25               0               0      <NA> 0.000000e+00     -1          2
## 26              44              45      MSTV 3.500000e-01      1       <NA>
## 27              46              47      MLTV 6.950000e+00      1       <NA>
## 28              48              49        UC 3.690524e-03      1       <NA>
## 29               0               0      <NA> 0.000000e+00     -1          2
## 30              50              51      MSTV 5.500000e-01      1       <NA>
## 31              52              53      ALTV 4.500000e+00      1       <NA>
## 32               0               0      <NA> 0.000000e+00     -1          2
## 33               0               0      <NA> 0.000000e+00     -1          3
## 34               0               0      <NA> 0.000000e+00     -1          3
## 35              54              55      Mode 1.220000e+02      1       <NA>
## 36               0               0      <NA> 0.000000e+00     -1          1
## 37               0               0      <NA> 0.000000e+00     -1          2
## 38              56              57    Median 1.545000e+02      1       <NA>
## 39              58              59       Min 9.400000e+01      1       <NA>
## 40               0               0      <NA> 0.000000e+00     -1          1
## 41              60              61      MLTV 1.210000e+01      1       <NA>
## 42              62              63       Max 1.595000e+02      1       <NA>
## 43               0               0      <NA> 0.000000e+00     -1          1
## 44              64              65        UC 4.542465e-03      1       <NA>
## 45              66              67        UC 4.549297e-03      1       <NA>
## 46              68              69      ASTV 8.050000e+01      1       <NA>
## 47               0               0      <NA> 0.000000e+00     -1          2
## 48               0               0      <NA> 0.000000e+00     -1          3
## 49               0               0      <NA> 0.000000e+00     -1          1
## 50              70              71      ALTV 7.500000e+00      1       <NA>
## 51              72              73       Min 1.440000e+02      1       <NA>
## 52               0               0      <NA> 0.000000e+00     -1          2
## 53               0               0      <NA> 0.000000e+00     -1          3
## 54              74              75      MLTV 2.250000e+00      1       <NA>
## 55              76              77      Nmax 1.000000e+01      1       <NA>
## 56               0               0      <NA> 0.000000e+00     -1          1
## 57              78              79        LB 1.430000e+02      1       <NA>
## 58               0               0      <NA> 0.000000e+00     -1          1
## 59              80              81       Max 1.550000e+02      1       <NA>
## 60               0               0      <NA> 0.000000e+00     -1          1
## 61               0               0      <NA> 0.000000e+00     -1          2
## 62              82              83      ALTV 5.950000e+01      1       <NA>
## 63               0               0      <NA> 0.000000e+00     -1          2
## 64              84              85        UC 1.101009e-03      1       <NA>
## 65               0               0      <NA> 0.000000e+00     -1          1
## 66              86              87     Width 7.000000e+01      1       <NA>
## 67              88              89        UC 7.059175e-03      1       <NA>
## 68              90              91      ALTV 6.300000e+01      1       <NA>
## 69               0               0      <NA> 0.000000e+00     -1          3
## 70              92              93     Width 7.800000e+01      1       <NA>
## 71              94              95      Mean 1.395000e+02      1       <NA>
## 72              96              97        DP 1.473432e-03      1       <NA>
```

```
## 73              0              0      <NA> 0.000000e+00     -1          2
## 74              0              0      <NA> 0.000000e+00     -1          3
## 75              0              0      <NA> 0.000000e+00     -1          2
## 76             98             99      MSTV 2.250000e+00      1       <NA>
## 77              0              0      <NA> 0.000000e+00     -1          2
## 78              0              0      <NA> 0.000000e+00     -1          2
## 79            100            101     Width 2.800000e+01      1       <NA>
## 80              0              0      <NA> 0.000000e+00     -1          2
## 81              0              0      <NA> 0.000000e+00     -1          1
## 82              0              0      <NA> 0.000000e+00     -1          1
## 83              0              0      <NA> 0.000000e+00     -1          2
## 84              0              0      <NA> 0.000000e+00     -1          3
## 85            102            103       Max 1.330000e+02      1       <NA>
## 86              0              0      <NA> 0.000000e+00     -1          2
## 87              0              0      <NA> 0.000000e+00     -1          3
## 88              0              0      <NA> 0.000000e+00     -1          1
## 89              0              0      <NA> 0.000000e+00     -1          3
## 90            104            105        AC 1.363310e-03      1       <NA>
## 91              0              0      <NA> 0.000000e+00     -1          3
## 92            106            107    Nzeros 5.000000e-01      1       <NA>
## 93              0              0      <NA> 0.000000e+00     -1          2
## 94            108            109       Min 9.900000e+01      1       <NA>
## 95            110            111       Min 1.355000e+02      1       <NA>
## 96            112            113        DL 1.172872e-02      1       <NA>
## 97            114            115       Min 6.200000e+01      1       <NA>
## 98              0              0      <NA> 0.000000e+00     -1          1
## 99            116            117       Min 5.200000e+01      1       <NA>
## 100           118            119        LB 1.470000e+02      1       <NA>
## 101             0              0      <NA> 0.000000e+00     -1          1
## 102             0              0      <NA> 0.000000e+00     -1          3
## 103             0              0      <NA> 0.000000e+00     -1          2
## 104             0              0      <NA> 0.000000e+00     -1          2
## 105             0              0      <NA> 0.000000e+00     -1          1
## 106             0              0      <NA> 0.000000e+00     -1          1
## 107           120            121      Mean 1.435000e+02      1       <NA>
## 108             0              0      <NA> 0.000000e+00     -1          2
## 109             0              0      <NA> 0.000000e+00     -1          1
## 110             0              0      <NA> 0.000000e+00     -1          2
## 111             0              0      <NA> 0.000000e+00     -1          1
## 112           122            123      Mean 1.485000e+02      1       <NA>
## 113           124            125      ASTV 5.850000e+01      1       <NA>
## 114             0              0      <NA> 0.000000e+00     -1          1
## 115           126            127        UC 6.175553e-03      1       <NA>
## 116             0              0      <NA> 0.000000e+00     -1          3
## 117             0              0      <NA> 0.000000e+00     -1          1
## 118           128            129      ALTV 2.650000e+01      1       <NA>
## 119             0              0      <NA> 0.000000e+00     -1          1
## 120             0              0      <NA> 0.000000e+00     -1          2
## 121             0              0      <NA> 0.000000e+00     -1          1
## 122           130            131      Mode 1.030000e+02      1       <NA>
```

```
## 123           132           133       Max 1.645000e+02      1       <NA>
## 124             0             0      <NA> 0.000000e+00     -1          1
## 125             0             0      <NA> 0.000000e+00     -1          3
## 126             0             0      <NA> 0.000000e+00     -1          2
## 127             0             0      <NA> 0.000000e+00     -1          3
## 128             0             0      <NA> 0.000000e+00     -1          1
## 129             0             0      <NA> 0.000000e+00     -1          1
## 130           134           135  Variance 1.050000e+02      1       <NA>
## 131           136           137       Max 1.365000e+02      1       <NA>
## 132           138           139       Min 9.100000e+01      1       <NA>
## 133           140           141       Min 1.390000e+02      1       <NA>
## 134             0             0      <NA> 0.000000e+00     -1          1
## 135             0             0      <NA> 0.000000e+00     -1          2
## 136           142           143        LB 1.225000e+02      1       <NA>
## 137           144           145      Mode 1.535000e+02      1       <NA>
## 138             0             0      <NA> 0.000000e+00     -1          2
## 139           146           147      MSTV 7.500000e-01      1       <NA>
## 140           148           149      MLTV 1.380000e+01      1       <NA>
## 141           150           151      Nmax 2.500000e+00      1       <NA>
## 142             0             0      <NA> 0.000000e+00     -1          1
## 143           152           153      Mode 1.240000e+02      1       <NA>
## 144           154           155        FM 3.216645e-01      1       <NA>
## 145           156           157  Variance 1.300000e+01      1       <NA>
## 146             0             0      <NA> 0.000000e+00     -1          1
## 147           158           159       Max 1.605000e+02      1       <NA>
## 148           160           161        AC 6.172840e-04      1       <NA>
## 149           162           163      MSTV 1.400000e+00      1       <NA>
## 150             0             0      <NA> 0.000000e+00     -1          1
## 151             0             0      <NA> 0.000000e+00     -1          2
## 152             0             0      <NA> 0.000000e+00     -1          1
## 153             0             0      <NA> 0.000000e+00     -1          2
## 154             0             0      <NA> 0.000000e+00     -1          1
## 155           164           165      MLTV 1.455000e+01      1       <NA>
## 156             0             0      <NA> 0.000000e+00     -1          2
## 157             0             0      <NA> 0.000000e+00     -1          1
## 158             0             0      <NA> 0.000000e+00     -1          1
## 159             0             0      <NA> 0.000000e+00     -1          2
## 160           166           167      ASTV 4.450000e+01      1       <NA>
## 161             0             0      <NA> 0.000000e+00     -1          1
## 162             0             0      <NA> 0.000000e+00     -1          2
## 163           168           169        UC 5.153023e-03      1       <NA>
## 164             0             0      <NA> 0.000000e+00     -1          1
## 165             0             0      <NA> 0.000000e+00     -1          2
## 166             0             0      <NA> 0.000000e+00     -1          2
## 167             0             0      <NA> 0.000000e+00     -1          1
## 168             0             0      <NA> 0.000000e+00     -1          1
## 169           170           171        UC 5.629126e-03      1       <NA>
## 170             0             0      <NA> 0.000000e+00     -1          2
## 171             0             0      <NA> 0.000000e+00     -1          1
```

Multi-Dimension scaling plot of proximity Matrix

```
MDSplot(rf, train$NSP)
```



Prediction and confusion Matrix

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.6.3
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
##
## Attaching package: 'ggplot2'
```

```
## The following object is masked from 'package:randomForest':
##
##     margin
```

```
prd <- predict(rf, train)
head(prd,20) #predicted
```

```
##  1  2  3  4  6  7  8  9 10 11 12 13 15 17 18 19 20 21 22 23
##  2  1  1  1  3  3  3  3  3  2  2  1  1  1  2  1  1  3  1  3
## Levels: 1 2 3
```

```
head(train$NSP, 20) #actual
```

```
##  [1] 2 1 1 1 3 3 3 3 3 2 2 1 1 1 2 1 1 3 1 3
## Levels: 1 2 3
```

prediction & Confusion Matrix on train data

```
confusionMatrix(prd, train$NSP)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    1    2    3
##          1 1195    1    0
##          2    0  200    0
##          3    0    0  127
##
## Overall Statistics
##
##                Accuracy : 0.9993
##                  95% CI : (0.9963, 1)
##     No Information Rate : 0.7846
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9982
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: 1 Class: 2 Class: 3
## Sensitivity            1.0000   0.9950  1.00000
## Specificity            0.9970   1.0000  1.00000
## Pos Pred Value         0.9992   1.0000  1.00000
## Neg Pred Value         1.0000   0.9992  1.00000
## Prevalence             0.7846   0.1320  0.08339
## Detection Rate         0.7846   0.1313  0.08339
## Detection Prevalence   0.7853   0.1313  0.08339
## Balanced Accuracy      0.9985   0.9975  1.00000
```

prediction & Confusion Matrix on test data

```
prd2 <- predict(rf, validate)
confusionMatrix(prd2, validate$NSP)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   1   2   3
##          1 457  19   5
##          2   2  73   2
```

```
##          3   1   2  42
##
## Overall Statistics
##
##                Accuracy : 0.9486
##                  95% CI : (0.9278, 0.9648)
##     No Information Rate : 0.7629
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8594
##
##  Mcnemar's Test P-Value : 0.0009261
##
## Statistics by Class:
##
##                      Class: 1 Class: 2 Class: 3
## Sensitivity            0.9935   0.7766  0.85714
## Specificity            0.8322   0.9921  0.99458
## Pos Pred Value         0.9501   0.9481  0.93333
## Neg Pred Value         0.9754   0.9601  0.98746
## Prevalence             0.7629   0.1559  0.08126
## Detection Rate         0.7579   0.1211  0.06965
## Detection Prevalence   0.7977   0.1277  0.07463
## Balanced Accuracy      0.9128   0.8844  0.92586
```