Machine Learning Practice

Load the data into R, run the following algorithms: logistic regressions, KNN, SVM, naïve bayes, decision trees, random forest and glmnet models.

```r
MyData<-read.table("ionosphere.data",sep = ",", stringsAsFactors = FALSE)
```

```r
Data<-na.omit(MyData)[,-c(1,2)]
Data$V35 <- factor(Data$V35, labels = c('0', '1'))
print("The 'g' is labeled as 0, 'g' is labeled as 1.")
```

```
## [1] "The 'g' is labeled as 0, 'g' is labeled as 1."
```

```r
table(Data$V35)
```

```
##
##   0   1
## 126 225
```

```r
#Create training dataset and test dataset
set.seed(1234)
index <- createDataPartition(Data$V35, p = 0.7, list = F)
traindata <- Data[index, ]
testdata <- Data[-index, ]
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: V35
##
## Terms added sequentially (first to last)
##
##
##       Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                   246     322.88
## V3     1   74.288       245     248.59 < 2.2e-16 ***
## V4     1    3.187       244     245.41 0.0742462 .
## V5     1   24.644       243     220.76 6.897e-07 ***
## V6     1    3.565       242     217.20 0.0590179 .
## V7     1    9.705       241     207.49 0.0018381 **
## V8     1   10.029       240     197.46 0.0015409 **
## V9     1    1.270       239     196.19 0.2598564
## V10    1    0.017       238     196.18 0.8970231
## V11    1    1.287       237     194.89 0.2566913
```

```
## V12   1    0.041        236    194.85 0.8403130
## V13   1    2.934        235    191.92 0.0867332 .
## V14   1    0.008        234    191.91 0.9266420
## V15   1    6.389        233    185.52 0.0114810 *
## V16   1    0.079        232    185.44 0.7789037
## V17   1    0.085        231    185.36 0.7707075
## V18   1    5.297        230    180.06 0.0213566 *
## V19   1    0.262        229    179.80 0.6090858
## V20   1    2.068        228    177.73 0.1503808
## V21   1   11.273        227    166.46 0.0007866 ***
## V22   1    4.312        226    162.14 0.0378341 *
## V23   1    1.081        225    161.06 0.2984946
## V24   1    3.642        224    157.42 0.0563236 .
## V25   1    0.621        223    156.80 0.4305042
## V26   1    5.255        222    151.54 0.0218860 *
## V27   1   25.968        221    125.57 3.471e-07 ***
## V28   1    4.140        220    121.44 0.0418892 *
## V29   1    1.703        219    119.73 0.1919324
## V30   1    0.974        218    118.76 0.3236103
## V31   1    0.052        217    118.71 0.8200788
## V32   1    3.591        216    115.12 0.0581039 .
## V33   1    0.090        215    115.03 0.7637194
## V34   1    4.501        214    110.52 0.0338701 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


##
## Call:
## glm(formula = V35 ~ ., family = binomial(link = "logit"), data = traindata)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.65313  -0.06812   0.14777   0.29866   2.92720
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.0658      0.7437  -4.122 3.75e-05 ***
## V3            2.0368      1.2907   1.578 0.114562
## V4            2.0959      1.1194   1.872 0.061159 .
## V5            3.8042      1.4648   2.597 0.009402 **
## V6            3.1946      1.1680   2.735 0.006237 **
## V7            1.2262      1.3932   0.880 0.378793
## V8            1.3865      1.1039   1.256 0.209116
## V9            1.8799      1.2723   1.478 0.139529
## V10          -0.6901      1.0723  -0.644 0.519820
## V11          -2.5435      1.3675  -1.860 0.062887 .
## V12          -1.1638      0.9302  -1.251 0.210879
## V13          -2.3851      1.4940  -1.596 0.110385
## V14           1.6099      0.9456   1.702 0.088669 .
## V15           2.4826      1.3468   1.843 0.065289 .
## V16           0.4868      1.1033   0.441 0.659063
## V17          -0.5604      1.1494  -0.488 0.625835
## V18           0.5052      0.8852   0.571 0.568162
## V19          -0.5546      1.1332  -0.489 0.624557
```

2

```
## V20            -1.1058      1.2331   -0.897 0.369830
## V21             0.4559      1.2200    0.374 0.708630
## V22            -3.8158      1.1785   -3.238 0.001204 **
## V23             3.8416      1.3059    2.942 0.003264 **
## V24             1.7956      0.8042    2.233 0.025565 *
## V25             1.6889      0.9197    1.836 0.066305 .
## V26             1.8063      1.0947    1.650 0.098949 .
## V27            -4.7915      1.2889   -3.718 0.000201 ***
## V28            -0.4645      1.0408   -0.446 0.655412
## V29             1.0159      1.0888    0.933 0.350794
## V30             1.9924      1.1097    1.795 0.072576 .
## V31            -0.7133      1.0527   -0.678 0.498009
## V32            -1.4854      1.1039   -1.346 0.178421
## V33             0.1913      1.2786    0.150 0.881045
## V34            -2.2674      1.1358   -1.996 0.045894 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 322.88  on 246  degrees of freedom
## Residual deviance: 110.52  on 214  degrees of freedom
## AIC: 176.52
##
## Number of Fisher Scoring iterations: 7


## Start:  AIC=176.52
## V35 ~ V3 + V4 + V5 + V6 + V7 + V8 + V9 + V10 + V11 + V12 + V13 +
##     V14 + V15 + V16 + V17 + V18 + V19 + V20 + V21 + V22 + V23 +
##     V24 + V25 + V26 + V27 + V28 + V29 + V30 + V31 + V32 + V33 +
##     V34
##
##        Df Deviance    AIC
## - V33   1   110.55 174.55
## - V21   1   110.66 174.66
## - V28   1   110.72 174.72
## - V16   1   110.72 174.72
## - V19   1   110.76 174.76
## - V17   1   110.76 174.76
## - V18   1   110.84 174.84
## - V10   1   110.96 174.96
## - V31   1   110.98 174.99
## - V7    1   111.33 175.33
## - V20   1   111.35 175.35
## - V29   1   111.39 175.39
## - V12   1   112.17 176.18
## - V8    1   112.26 176.26
## - V32   1   112.43 176.43
## <none>      110.52 176.52
## - V9    1   112.59 176.59
## - V13   1   113.14 177.14
## - V3    1   113.49 177.49
## - V26   1   113.62 177.62
## - V14   1   113.63 177.63
```

```
## - V15    1    113.69 177.69
## - V25    1    113.89 177.89
## - V11    1    114.09 178.09
## - V30    1    114.30 178.30
## - V4     1    114.48 178.48
## - V34    1    115.03 179.03
## - V24    1    115.96 179.96
## - V5     1    118.03 182.03
## - V6     1    118.26 182.26
## - V23    1    120.04 184.04
## - V22    1    123.51 187.51
## - V27    1    136.86 200.86
##
## Step:  AIC=174.55
## V35 ~ V3 + V4 + V5 + V6 + V7 + V8 + V9 + V10 + V11 + V12 + V13 +
##     V14 + V15 + V16 + V17 + V18 + V19 + V20 + V21 + V22 + V23 +
##     V24 + V25 + V26 + V27 + V28 + V29 + V30 + V31 + V32 + V34
##
##          Df Deviance    AIC
## - V16    1    110.72 172.72
## - V21    1    110.73 172.73
## - V19    1    110.77 172.77
## - V28    1    110.83 172.82
## - V17    1    110.83 172.83
## - V18    1    110.86 172.86
## - V10    1    110.96 172.96
## - V31    1    110.99 172.99
## - V20    1    111.36 173.36
## - V7     1    111.40 173.40
## - V29    1    111.53 173.53
## - V12    1    112.17 174.18
## - V8     1    112.29 174.29
## - V32    1    112.47 174.47
## <none>        110.55 174.55
## - V9     1    112.89 174.88
## - V13    1    113.38 175.38
## - V3     1    113.50 175.50
## - V14    1    113.63 175.63
## - V26    1    113.88 175.88
## - V25    1    113.90 175.90
## - V15    1    113.93 175.93
## - V11    1    114.10 176.10
## + V33    1    110.52 176.52
## - V4     1    114.64 176.64
## - V30    1    114.68 176.68
## - V34    1    115.12 177.12
## - V24    1    116.12 178.12
## - V5     1    118.08 180.08
## - V6     1    118.31 180.31
## - V23    1    120.09 182.09
## - V22    1    123.61 185.61
## - V27    1    139.29 201.29
##
## Step:  AIC=172.72
```

4

```
## V35 ~ V3 + V4 + V5 + V6 + V7 + V8 + V9 + V10 + V11 + V12 + V13 +
##     V14 + V15 + V17 + V18 + V19 + V20 + V21 + V22 + V23 + V24 +
##     V25 + V26 + V27 + V28 + V29 + V30 + V31 + V32 + V34
##
##          Df Deviance    AIC
## - V21   1   110.88 170.88
## - V17   1   110.93 170.93
## - V19   1   110.96 170.96
## - V10   1   111.08 171.08
## - V31   1   111.15 171.15
## - V18   1   111.15 171.15
## - V28   1   111.17 171.17
## - V7    1   111.47 171.47
## - V20   1   111.67 171.67
## - V29   1   111.79 171.79
## - V12   1   112.22 172.22
## - V32   1   112.65 172.65
## <none>      110.72 172.72
## - V9    1   113.18 173.18
## - V8    1   113.36 173.36
## - V13   1   113.55 173.55
## - V3    1   113.88 173.88
## - V14   1   113.90 173.90
## - V25   1   113.97 173.97
## + V16   1   110.55 174.55
## - V15   1   114.61 174.61
## - V11   1   114.64 174.64
## + V33   1   110.72 174.72
## - V30   1   114.72 174.72
## - V4    1   114.76 174.76
## - V34   1   115.52 175.52
## - V24   1   116.17 176.17
## - V26   1   116.38 176.38
## - V6    1   118.36 178.36
## - V5    1   118.44 178.44
## - V23   1   120.39 180.39
## - V22   1   124.11 184.11
## - V27   1   139.81 199.81
##
## Step:  AIC=170.88
## V35 ~ V3 + V4 + V5 + V6 + V7 + V8 + V9 + V10 + V11 + V12 + V13 +
##     V14 + V15 + V17 + V18 + V19 + V20 + V22 + V23 + V24 + V25 +
##     V26 + V27 + V28 + V29 + V30 + V31 + V32 + V34
##
##          Df Deviance    AIC
## - V17   1   111.03 169.03
## - V19   1   111.07 169.07
## - V10   1   111.19 169.19
## - V18   1   111.26 169.26
## - V31   1   111.34 169.34
## - V28   1   111.35 169.35
## - V7    1   111.64 169.64
## - V20   1   111.86 169.86
## - V29   1   112.14 170.14
```

```
## - V12    1   112.49 170.49
## - V32    1   112.69 170.69
## <none>       110.88 170.88
## - V13    1   113.55 171.55
## - V8     1   113.64 171.64
## - V9     1   113.71 171.71
## - V3     1   113.89 171.88
## - V14    1   114.51 172.51
## - V25    1   114.51 172.51
## + V21    1   110.72 172.72
## + V16    1   110.73 172.73
## + V33    1   110.84 172.84
## - V11    1   114.90 172.90
## - V30    1   114.98 172.98
## - V4     1   114.98 172.98
## - V15    1   115.20 173.20
## - V34    1   115.69 173.69
## - V24    1   116.29 174.29
## - V26    1   117.50 175.50
## - V5     1   118.56 176.56
## - V6     1   118.63 176.63
## - V23    1   121.11 179.11
## - V22    1   128.33 186.33
## - V27    1   140.84 198.84
##
## Step:  AIC=169.03
## V35 ~ V3 + V4 + V5 + V6 + V7 + V8 + V9 + V10 + V11 + V12 + V13 +
##     V14 + V15 + V18 + V19 + V20 + V22 + V23 + V24 + V25 + V26 +
##     V27 + V28 + V29 + V30 + V31 + V32 + V34
##
##          Df Deviance    AIC
## - V19    1   111.28 167.28
## - V10    1   111.33 167.33
## - V18    1   111.35 167.35
## - V28    1   111.63 167.63
## - V31    1   111.69 167.69
## - V7     1   111.80 167.80
## - V20    1   112.00 168.00
## - V12    1   112.54 168.54
## - V29    1   112.63 168.63
## - V32    1   112.80 168.80
## <none>       111.03 169.03
## - V9     1   113.75 169.75
## - V13    1   113.75 169.75
## - V3     1   114.21 170.21
## - V8     1   114.31 170.31
## - V25    1   114.57 170.57
## - V14    1   114.78 170.78
## + V17    1   110.88 170.88
## + V21    1   110.93 170.93
## + V16    1   110.95 170.95
## + V33    1   110.97 170.97
## - V4     1   114.99 170.99
## - V30    1   115.09 171.09
```

```
## - V11    1    115.20 171.21
## - V15    1    115.31 171.31
## - V34    1    116.13 172.13
## - V24    1    116.85 172.85
## - V26    1    118.89 174.89
## - V6     1    118.91 174.91
## - V5     1    119.20 175.20
## - V23    1    121.83 177.83
## - V22    1    128.75 184.75
## - V27    1    141.25 197.25
##
## Step:  AIC=167.28
## V35 ~ V3 + V4 + V5 + V6 + V7 + V8 + V9 + V10 + V11 + V12 + V13 +
##      V14 + V15 + V18 + V20 + V22 + V23 + V24 + V25 + V26 + V27 +
##      V28 + V29 + V30 + V31 + V32 + V34
##
##           Df Deviance    AIC
## - V10    1    111.46 165.46
## - V18    1    111.57 165.57
## - V28    1    111.75 165.75
## - V7     1    111.99 165.99
## - V31    1    112.25 166.25
## - V12    1    112.77 166.77
## - V29    1    112.85 166.85
## - V20    1    112.95 166.95
## - V32    1    113.02 167.02
## <none>        111.28 167.28
## - V13    1    113.90 167.90
## - V9     1    114.08 168.08
## - V3     1    114.21 168.21
## - V8     1    114.64 168.64
## - V25    1    114.66 168.66
## + V19    1    111.03 169.03
## + V17    1    111.07 169.07
## + V16    1    111.18 169.18
## + V21    1    111.22 169.22
## + V33    1    111.25 169.25
## - V15    1    115.56 169.56
## - V11    1    115.66 169.66
## - V14    1    115.70 169.70
## - V30    1    115.75 169.75
## - V4     1    116.27 170.27
## - V34    1    116.59 170.59
## - V24    1    116.85 170.85
## - V26    1    118.89 172.90
## - V6     1    119.10 173.10
## - V5     1    120.66 174.66
## - V23    1    121.84 175.84
## - V22    1    128.75 182.75
## - V27    1    143.45 197.45
##
## Step:  AIC=165.46
## V35 ~ V3 + V4 + V5 + V6 + V7 + V8 + V9 + V11 + V12 + V13 + V14 +
##      V15 + V18 + V20 + V22 + V23 + V24 + V25 + V26 + V27 + V28 +
```

```
##       V29 + V30 + V31 + V32 + V34
##
##          Df Deviance    AIC
## - V18    1    111.67 163.67
## - V28    1    111.98 163.98
## - V7     1    112.06 164.06
## - V31    1    112.48 164.48
## - V20    1    113.03 165.03
## - V32    1    113.05 165.05
## - V29    1    113.37 165.37
## - V12    1    113.39 165.39
## <none>        111.46 165.46
## - V13    1    113.92 165.92
## - V3     1    114.21 166.21
## - V25    1    114.69 166.69
## - V8     1    114.75 166.75
## - V9     1    114.90 166.90
## + V17    1    111.28 167.28
## + V10    1    111.28 167.28
## + V19    1    111.33 167.33
## + V16    1    111.39 167.39
## + V21    1    111.40 167.40
## + V33    1    111.45 167.45
## - V15    1    115.64 167.64
## - V11    1    115.69 167.69
## - V30    1    115.75 167.75
## - V14    1    116.59 168.59
## - V34    1    116.75 168.75
## - V4     1    117.00 169.00
## - V24    1    117.01 169.01
## - V6     1    119.10 171.10
## - V26    1    120.36 172.36
## - V5     1    121.11 173.11
## - V23    1    121.96 173.96
## - V22    1    130.86 182.86
## - V27    1    143.46 195.46
##
## Step:  AIC=163.67
## V35 ~ V3 + V4 + V5 + V6 + V7 + V8 + V9 + V11 + V12 + V13 + V14 +
##     V15 + V20 + V22 + V23 + V24 + V25 + V26 + V27 + V28 + V29 +
##     V30 + V31 + V32 + V34
##
##          Df Deviance    AIC
## - V28    1    112.01 162.01
## - V7     1    112.53 162.53
## - V31    1    112.61 162.61
## - V20    1    113.04 163.04
## - V32    1    113.08 163.08
## - V29    1    113.47 163.47
## - V12    1    113.59 163.59
## <none>        111.67 163.67
## - V3     1    114.39 164.39
## - V13    1    114.43 164.43
## - V25    1    114.78 164.78
```

```
## - V8     1   114.93 164.93
## - V9     1   115.13 165.13
## + V18    1   111.46 165.46
## + V16    1   111.53 165.53
## + V19    1   111.54 165.54
## + V17    1   111.55 165.55
## + V10    1   111.57 165.57
## + V21    1   111.62 165.62
## + V33    1   111.67 165.67
## - V11    1   115.69 165.69
## - V15    1   115.71 165.71
## - V30    1   115.88 165.88
## - V14    1   116.77 166.77
## - V24    1   117.47 167.47
## - V4     1   117.59 167.59
## - V34    1   118.49 168.49
## - V6     1   119.81 169.81
## - V26    1   120.36 170.36
## - V5     1   121.13 171.13
## - V23    1   122.08 172.08
## - V22    1   131.33 181.33
## - V27    1   143.46 193.46
##
## Step:  AIC=162.01
## V35 ~ V3 + V4 + V5 + V6 + V7 + V8 + V9 + V11 + V12 + V13 + V14 +
##      V15 + V20 + V22 + V23 + V24 + V25 + V26 + V27 + V29 + V30 +
##      V31 + V32 + V34
##
##          Df Deviance    AIC
## - V31    1   112.88 160.88
## - V7     1   112.91 160.91
## - V20    1   113.40 161.40
## - V12    1   113.99 161.99
## <none>       112.01 162.01
## - V29    1   114.02 162.02
## - V32    1   114.17 162.17
## - V13    1   114.67 162.66
## - V3     1   114.75 162.75
## - V25    1   114.86 162.86
## - V8     1   115.07 163.07
## - V9     1   115.34 163.34
## + V28    1   111.67 163.67
## - V11    1   115.70 163.70
## + V16    1   111.78 163.78
## + V17    1   111.80 163.80
## - V15    1   115.83 163.83
## + V10    1   111.83 163.84
## + V21    1   111.93 163.93
## + V33    1   111.94 163.94
## + V19    1   111.96 163.96
## + V18    1   111.98 163.98
## - V30    1   116.27 164.27
## - V24    1   117.51 165.51
## - V4     1   118.03 166.03
```

```
## - V14    1   118.39 166.39
## - V34    1   119.76 167.76
## - V26    1   120.39 168.39
## - V6     1   120.39 168.39
## - V5     1   121.24 169.24
## - V23    1   123.00 171.00
## - V22    1   134.27 182.27
## - V27    1   143.69 191.69
##
## Step:  AIC=160.88
## V35 ~ V3 + V4 + V5 + V6 + V7 + V8 + V9 + V11 + V12 + V13 + V14 +
##     V15 + V20 + V22 + V23 + V24 + V25 + V26 + V27 + V29 + V30 +
##     V32 + V34
##
##          Df Deviance    AIC
## - V7     1   113.45 159.45
## - V20    1   113.72 159.72
## - V29    1   114.38 160.38
## - V32    1   114.71 160.71
## - V12    1   114.76 160.76
## <none>       112.88 160.88
## - V13    1   115.09 161.09
## - V25    1   115.26 161.26
## - V3     1   115.72 161.72
## - V8     1   115.99 161.99
## + V31    1   112.01 162.01
## - V11    1   116.08 162.08
## - V15    1   116.16 162.16
## - V30    1   116.35 162.35
## - V9     1   116.37 162.37
## + V17    1   112.42 162.42
## + V28    1   112.61 162.61
## + V10    1   112.66 162.66
## + V16    1   112.71 162.71
## + V19    1   112.74 162.74
## + V21    1   112.78 162.78
## + V18    1   112.87 162.87
## + V33    1   112.88 162.88
## - V24    1   118.61 164.61
## - V14    1   118.70 164.70
## - V4     1   118.78 164.78
## - V34    1   120.04 166.04
## - V6     1   120.59 166.59
## - V5     1   121.25 167.25
## - V26    1   121.90 167.90
## - V23    1   123.03 169.03
## - V22    1   134.41 180.41
## - V27    1   144.22 190.22
##
## Step:  AIC=159.45
## V35 ~ V3 + V4 + V5 + V6 + V8 + V9 + V11 + V12 + V13 + V14 + V15 +
##     V20 + V22 + V23 + V24 + V25 + V26 + V27 + V29 + V30 + V32 +
##     V34
##
```

```
##           Df Deviance    AIC
## - V20     1   114.11 158.11
## - V32     1   115.02 159.02
## - V29     1   115.28 159.28
## - V13     1   115.39 159.39
## <none>        113.45 159.45
## - V12     1   115.58 159.58
## - V25     1   115.67 159.67
## - V8      1   116.44 160.44
## - V30     1   116.66 160.66
## + V7      1   112.88 160.88
## + V31     1   112.91 160.91
## + V17     1   113.05 161.05
## + V28     1   113.09 161.09
## - V3      1   117.16 161.16
## + V16     1   113.28 161.28
## + V21     1   113.33 161.33
## + V19     1   113.34 161.34
## + V18     1   113.37 161.37
## + V33     1   113.40 161.40
## + V10     1   113.41 161.41
## - V11     1   117.99 161.99
## - V9      1   117.99 161.99
## - V15     1   118.37 162.37
## - V4      1   119.60 163.60
## - V24     1   119.64 163.63
## - V34     1   120.99 164.99
## - V14     1   121.06 165.06
## - V6      1   121.42 165.42
## - V26     1   122.16 166.16
## - V23     1   123.27 167.26
## - V5      1   125.38 169.38
## - V22     1   135.57 179.57
## - V27     1   144.74 188.74
##
## Step:  AIC=158.11
## V35 ~ V3 + V4 + V5 + V6 + V8 + V9 + V11 + V12 + V13 + V14 + V15 +
##     V22 + V23 + V24 + V25 + V26 + V27 + V29 + V30 + V32 + V34
##
##           Df Deviance    AIC
## - V32     1   115.65 157.65
## - V12     1   115.89 157.89
## <none>        114.11 158.11
## - V13     1   116.33 158.33
## - V25     1   116.37 158.37
## - V30     1   116.66 158.66
## - V8      1   116.80 158.79
## - V29     1   116.86 158.87
## - V3      1   117.24 159.24
## + V20     1   113.45 159.45
## + V17     1   113.64 159.64
## + V7      1   113.72 159.72
## + V28     1   113.76 159.76
## + V19     1   113.77 159.77
```

```
## + V16    1   113.84 159.84
## + V31    1   113.88 159.88
## + V21    1   113.94 159.94
## + V10    1   114.08 160.08
## + V33    1   114.11 160.11
## + V18    1   114.11 160.11
## - V15    1   118.40 160.40
## - V11    1   118.60 160.60
## - V9     1   119.39 161.39
## - V4     1   119.77 161.77
## - V14    1   121.09 163.09
## - V24    1   121.22 163.22
## - V6     1   121.52 163.52
## - V34    1   121.91 163.91
## - V26    1   122.34 164.34
## - V23    1   125.13 167.13
## - V5     1   126.42 168.43
## - V22    1   139.09 181.09
## - V27    1   145.31 187.31
##
## Step:  AIC=157.65
## V35 ~ V3 + V4 + V5 + V6 + V8 + V9 + V11 + V12 + V13 + V14 + V15 +
##       V22 + V23 + V24 + V25 + V26 + V27 + V29 + V30 + V34
##
##          Df Deviance    AIC
## - V25    1   117.25 157.25
## <none>       115.65 157.65
## - V12    1   117.93 157.93
## - V13    1   118.05 158.05
## - V3     1   118.08 158.08
## + V32    1   114.11 158.11
## - V30    1   118.17 158.17
## - V8     1   118.29 158.29
## + V28    1   114.64 158.64
## - V29    1   118.68 158.68
## + V20    1   115.02 159.02
## + V16    1   115.24 159.24
## + V17    1   115.28 159.28
## + V19    1   115.33 159.32
## + V7     1   115.40 159.40
## + V18    1   115.42 159.43
## + V31    1   115.56 159.56
## + V21    1   115.61 159.60
## + V33    1   115.63 159.63
## + V10    1   115.65 159.65
## - V15    1   120.79 160.79
## - V9     1   121.48 161.48
## - V24    1   121.52 161.52
## - V6     1   121.67 161.68
## - V4     1   121.81 161.81
## - V11    1   121.97 161.97
## - V26    1   122.36 162.35
## - V14    1   124.86 164.86
## - V34    1   126.26 166.26
```

12

```
## - V23    1    127.63 167.63
## - V5     1    130.17 170.17
## - V22    1    139.46 179.46
## - V27    1    148.84 188.84
##
## Step:  AIC=157.25
## V35 ~ V3 + V4 + V5 + V6 + V8 + V9 + V11 + V12 + V13 + V14 + V15 +
##     V22 + V23 + V24 + V26 + V27 + V29 + V30 + V34
##
##         Df Deviance    AIC
## - V13    1    118.81 156.81
## <none>       117.25 157.25
## + V25    1    115.65 157.65
## - V30    1    119.70 157.70
## - V3     1    119.90 157.90
## - V12    1    120.03 158.03
## + V32    1    116.37 158.37
## + V20    1    116.57 158.57
## + V28    1    116.87 158.87
## + V19    1    116.96 158.96
## + V21    1    116.97 158.97
## + V7     1    117.06 159.06
## + V18    1    117.08 159.08
## + V17    1    117.09 159.09
## + V33    1    117.22 159.22
## + V31    1    117.23 159.23
## + V16    1    117.24 159.24
## + V10    1    117.25 159.25
## - V8     1    121.29 159.29
## - V29    1    121.92 159.93
## - V4     1    122.71 160.71
## - V6     1    122.75 160.75
## - V15    1    123.42 161.42
## - V24    1    123.94 161.94
## - V9     1    124.06 162.06
## - V11    1    124.74 162.74
## - V26    1    125.42 163.42
## - V14    1    127.30 165.30
## - V34    1    128.04 166.04
## - V23    1    129.01 167.01
## - V5     1    131.64 169.64
## - V22    1    144.97 182.97
## - V27    1    149.11 187.11
##
## Step:  AIC=156.81
## V35 ~ V3 + V4 + V5 + V6 + V8 + V9 + V11 + V12 + V14 + V15 + V22 +
##     V23 + V24 + V26 + V27 + V29 + V30 + V34
##
##         Df Deviance    AIC
## <none>       118.81 156.81
## + V13    1    117.25 157.25
## - V12    1    121.47 157.47
## + V32    1    117.70 157.70
## + V20    1    117.89 157.89
```

```
## + V25   1   118.05 158.05
## - V30   1   122.23 158.23
## - V29   1   122.34 158.34
## + V19   1   118.44 158.44
## + V28   1   118.53 158.53
## + V17   1   118.55 158.55
## + V16   1   118.62 158.62
## + V18   1   118.71 158.71
## + V10   1   118.77 158.77
## + V21   1   118.78 158.78
## + V7    1   118.79 158.79
## + V33   1   118.80 158.80
## + V31   1   118.81 158.81
## - V4    1   123.27 159.27
## - V15   1   123.48 159.48
## - V6    1   124.14 160.14
## - V8    1   124.79 160.79
## - V3    1   124.99 160.99
## - V9    1   125.79 161.79
## - V24   1   125.85 161.85
## - V26   1   127.58 163.58
## - V11   1   129.18 165.18
## - V5    1   131.64 167.64
## - V23   1   132.78 168.78
## - V34   1   133.01 169.01
## - V14   1   135.20 171.20
## - V27   1   151.28 187.28
## - V22   1   155.01 191.01


##
## Call:
## glm(formula = V35 ~ V3 + V4 + V5 + V6 + V8 + V9 + V11 + V12 +
##     V14 + V15 + V22 + V23 + V24 + V26 + V27 + V29 + V30 + V34,
##     family = binomial(link = "logit"), data = traindata)
##
## Deviance Residuals:
##     Min       1Q    Median       3Q      Max
## -2.6355  -0.1038   0.1621   0.3250   2.9863
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.9334     0.6112  -4.800 1.59e-06 ***
## V3            2.0574     0.9539   2.157 0.031009 *
## V4            1.7868     0.8649   2.066 0.038828 *
## V5            3.5701     1.0519   3.394 0.000689 ***
## V6            1.8616     0.8092   2.300 0.021427 *
## V8            2.0742     0.8960   2.315 0.020613 *
## V9            2.2526     0.8613   2.615 0.008916 **
## V11          -3.2241     1.0513  -3.067 0.002164 **
## V12          -1.1785     0.7597  -1.551 0.120866
## V14           2.5082     0.6892   3.639 0.000273 ***
## V15           1.8439     0.8734   2.111 0.034754 *
## V22          -4.5876     0.9304  -4.931 8.19e-07 ***
## V23           3.3210     0.9606   3.457 0.000546 ***
```

```
## V24           1.6587     0.6386    2.597 0.009395 **
## V26           1.7930     0.6526    2.748 0.006002 **
## V27          -4.3903     0.9787   -4.486 7.27e-06 ***
## V29           1.3967     0.7592    1.840 0.065806 .
## V30           1.3123     0.7317    1.793 0.072903 .
## V34          -2.8994     0.8566   -3.385 0.000712 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 322.88  on 246  degrees of freedom
## Residual deviance: 118.81  on 228  degrees of freedom
## AIC: 156.81
##
## Number of Fisher Scoring iterations: 7


## [1] 156.8074


## [1] 176.524


## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 28  3
##          1  9 64
##
##                Accuracy : 0.8846
##                  95% CI : (0.8071, 0.9389)
##     No Information Rate : 0.6442
##     P-Value [Acc > NIR] : 2.487e-08
##
##                   Kappa : 0.7388
##
##  Mcnemar's Test P-Value : 0.1489
##
##             Sensitivity : 0.7568
##             Specificity : 0.9552
##          Pos Pred Value : 0.9032
##          Neg Pred Value : 0.8767
##              Prevalence : 0.3558
##          Detection Rate : 0.2692
##    Detection Prevalence : 0.2981
##       Balanced Accuracy : 0.8560
##
##        'Positive' Class : 0
##


## [1] 88.46154


## Warning in roc.default(testdata$V33, glm_predict): 'response' has more than two
## levels. Consider setting 'levels' explicitly or using 'multiclass.roc' instead
```

```
## Setting levels: control = -1, case = -0.8121
```

```
## Setting direction: controls < cases
```
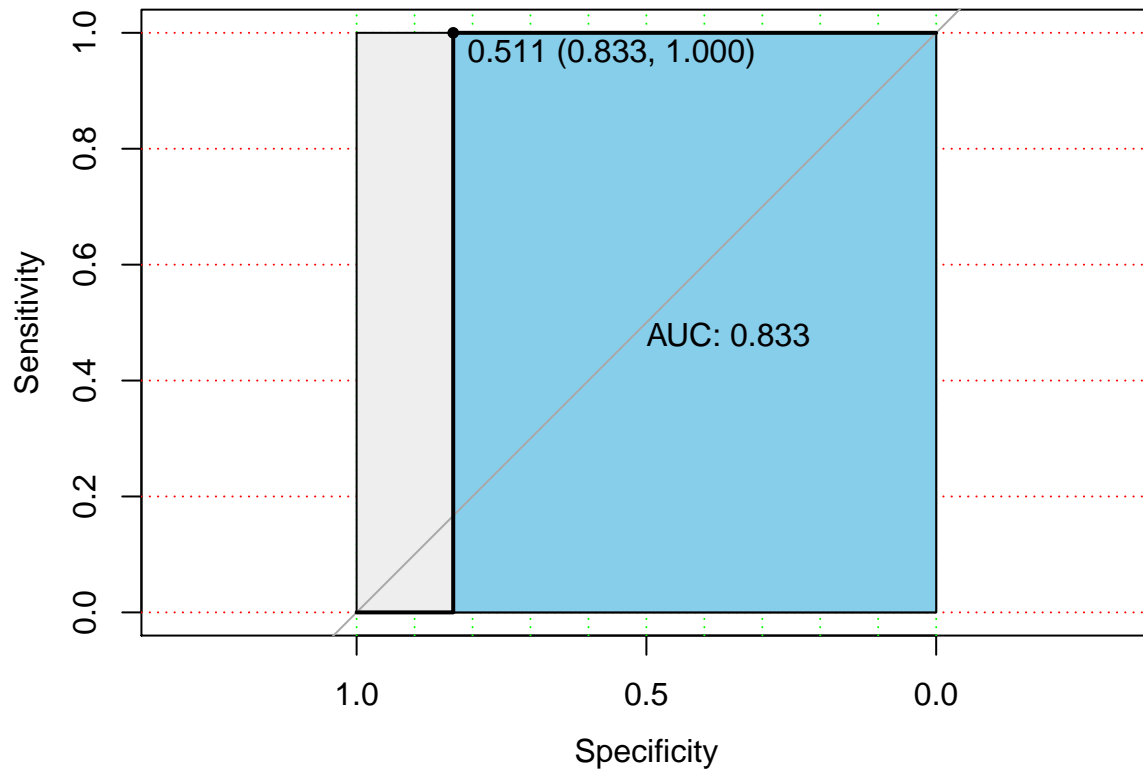


```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 23 14
##          1  2 65
##
##               Accuracy : 0.8462
##                 95% CI : (0.7622, 0.9094)
##    No Information Rate : 0.7596
##    P-Value [Acc > NIR] : 0.02171
##
##                  Kappa : 0.6381
##
##  Mcnemar's Test P-Value : 0.00596
##
##            Sensitivity : 0.9200
##            Specificity : 0.8228
##         Pos Pred Value : 0.6216
##         Neg Pred Value : 0.9701
##             Prevalence : 0.2404
```

```
##           Detection Rate : 0.2212
##     Detection Prevalence : 0.3558
##        Balanced Accuracy : 0.8714
##
##         'Positive' Class : 0
##
```

Error rate graph
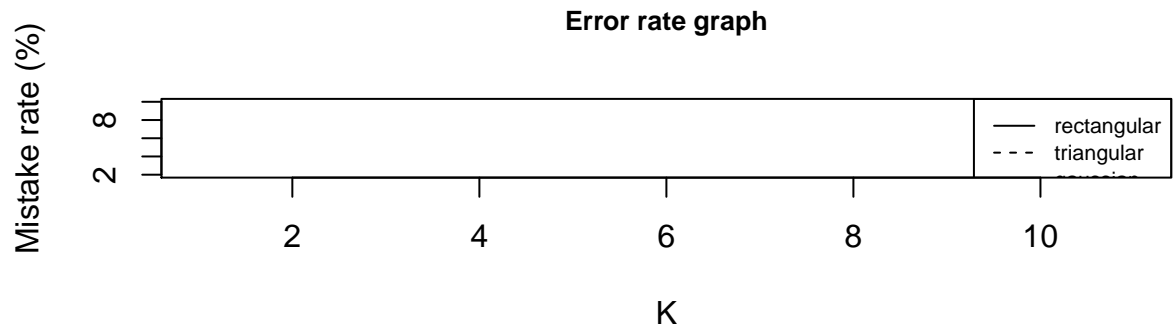

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0   1
##          0 33   2
##          1  4  65
##
##                 Accuracy : 0.9423
##                   95% CI : (0.8787, 0.9785)
##      No Information Rate : 0.6442
##      P-Value [Acc > NIR] : 6.661e-13
##
##                    Kappa : 0.8726
##
##   Mcnemar's Test P-Value : 0.6831
##
##              Sensitivity : 0.8919
##              Specificity : 0.9701
##           Pos Pred Value : 0.9429
```

```
##          Neg Pred Value : 0.9420
##             Prevalence : 0.3558
##         Detection Rate : 0.3173
##   Detection Prevalence : 0.3365
##      Balanced Accuracy : 0.9310
##
##       'Positive' Class : 0
##


## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 32 14
##          1  5 53
##
##               Accuracy : 0.8173
##                 95% CI : (0.7295, 0.8863)
##    No Information Rate : 0.6442
##    P-Value [Acc > NIR] : 8.509e-05
##
##                  Kappa : 0.622
##
##  Mcnemar's Test P-Value : 0.06646
##
##            Sensitivity : 0.8649
##            Specificity : 0.7910
##         Pos Pred Value : 0.6957
##         Neg Pred Value : 0.9138
##             Prevalence : 0.3558
##         Detection Rate : 0.3077
##   Detection Prevalence : 0.4423
##      Balanced Accuracy : 0.8280
##
##       'Positive' Class : 0
##


##
## Classification tree:
## rpart(formula = V35 ~ ., data = traindata, method = "class",
##     parms = list(split = "gini"))
##
## Variables actually used in tree construction:
## [1] V27 V5
##
## Root node error: 89/247 = 0.36032
##
## n= 247
##
##        CP nsplit rel error  xerror     xstd
## 1 0.53933      0   1.00000 1.00000 0.084778
## 2 0.20225      1   0.46067 0.55056 0.070419
## 3 0.01000      2   0.25843 0.30337 0.055100
```
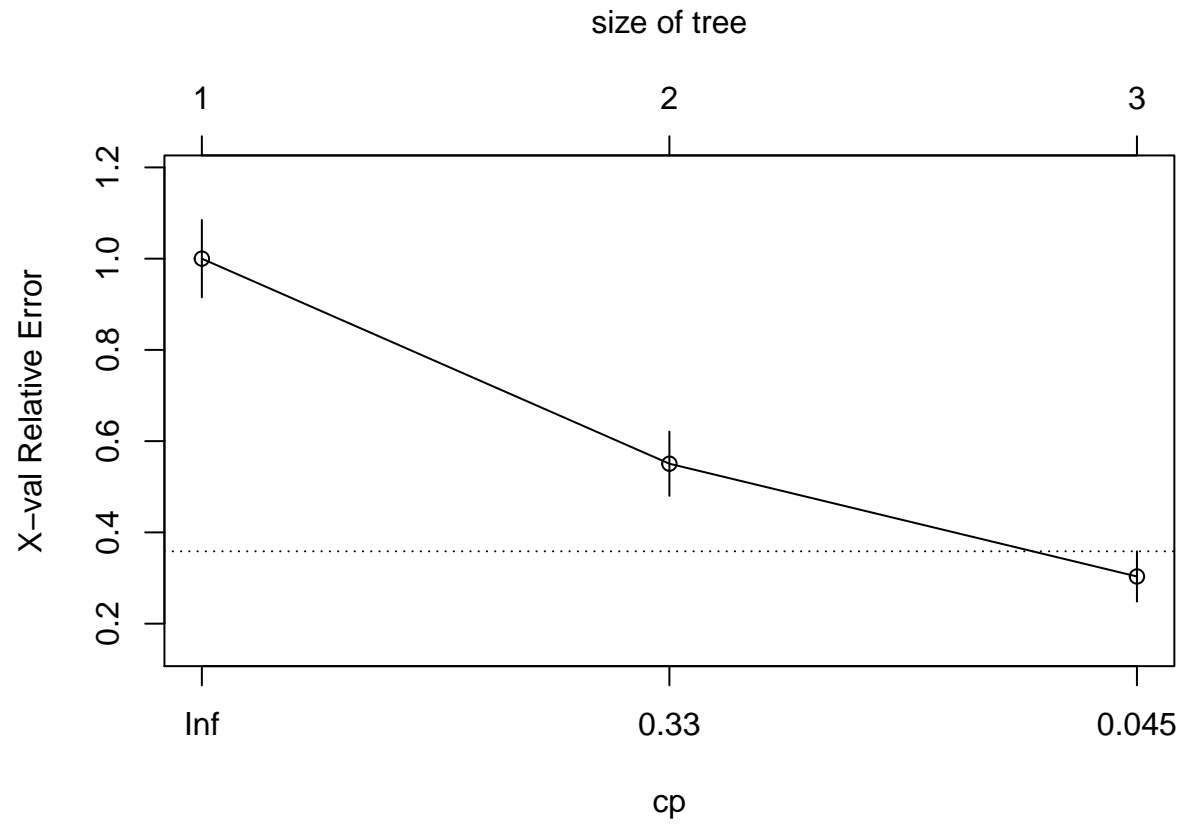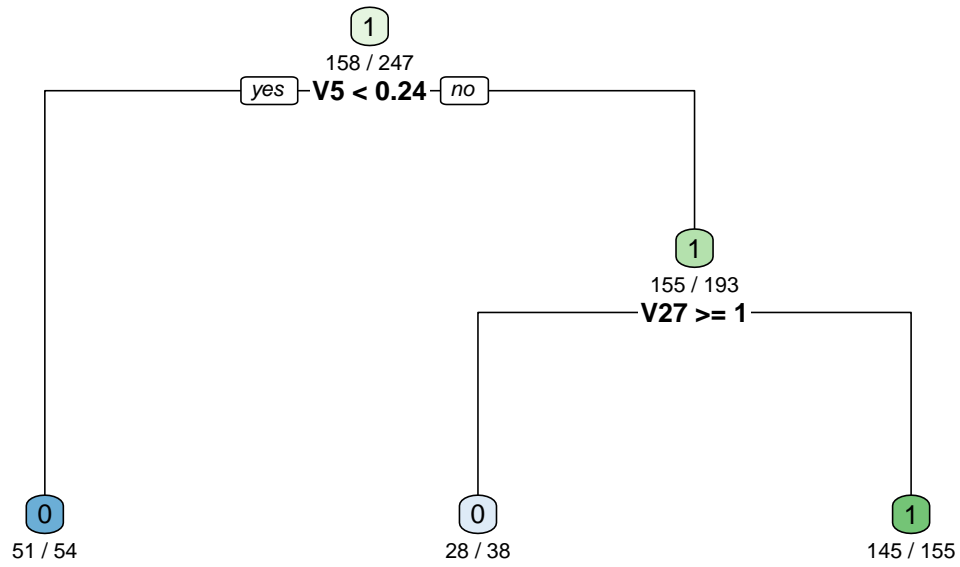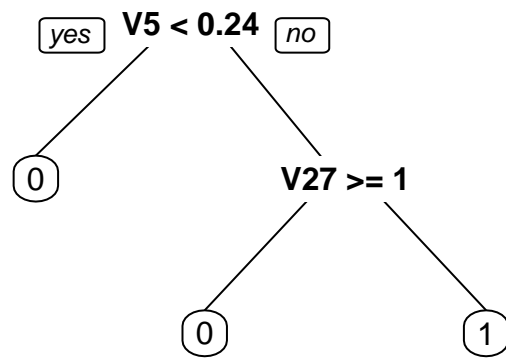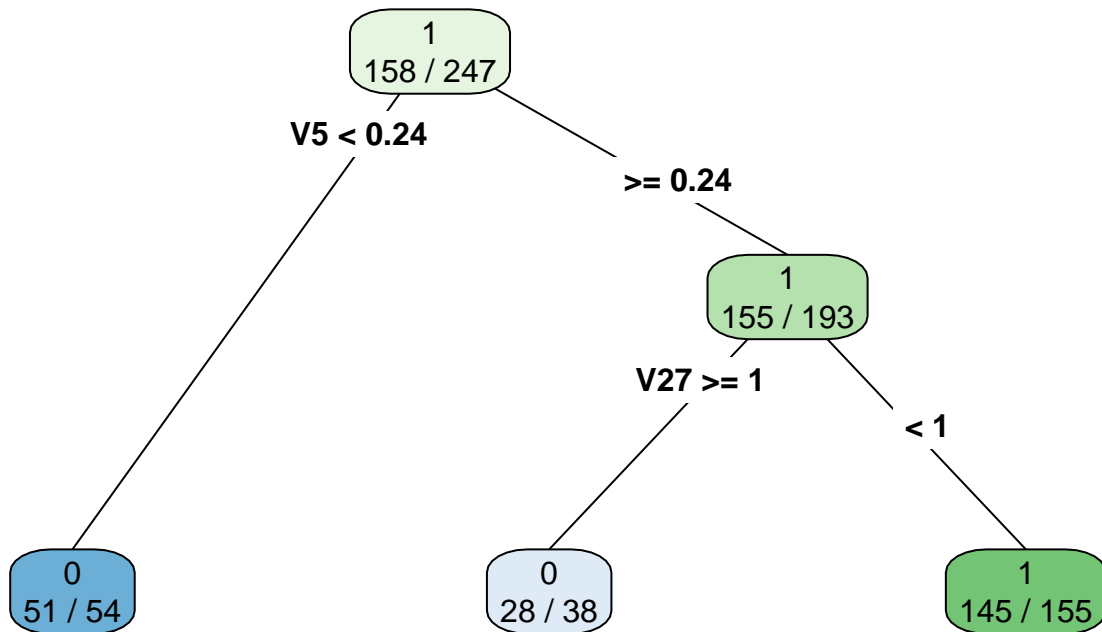
Visualization of CART

```
                              ┌─┐
                              │1│
                              └─┘
                           158 / 247
              ┌─────────┤yes├─ V5 < 0.24 ─┤no├─────────┐
              │                                        │
              │                                      ┌─┐
              │                                      │1│
              │                                      └─┘
              │                                   155 / 193
              │                        ┌────────── V27 >= 1 ──────────┐
              │                        │                              │
            ┌─┐                      ┌─┐                            ┌─┐
            │0│                      │0│                            │1│
            └─┘                      └─┘                            └─┘
          51 / 54                  28 / 38                       145 / 155
```

yes **V5 < 0.24** no

0

V27 >= 1

0

1

Visualization of CART


```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 33  6
##          1  4 61
##
##                Accuracy : 0.9038
##                  95% CI : (0.8303, 0.9529)
##     No Information Rate : 0.6442
##     P-Value [Acc > NIR] : 1.168e-09
##
##                   Kappa : 0.7927
##
##  Mcnemar's Test P-Value : 0.7518
##
##             Sensitivity : 0.8919
##             Specificity : 0.9104
##          Pos Pred Value : 0.8462
##          Neg Pred Value : 0.9385
##              Prevalence : 0.3558
##          Detection Rate : 0.3173
##    Detection Prevalence : 0.3750
##       Balanced Accuracy : 0.9012
##
##        'Positive' Class : 0
```

```
## 

##             0           1
## 2 0.4723618 0.52763819
## 4 0.5056180 0.49438202
## 5 0.3314607 0.66853933
## 6 0.8870056 0.11299435
## 7 0.1235294 0.87647059
## 8 0.9058824 0.09411765


## [1] 199 178 178 177 170 170


## [1] 17 15 16 17 17 18


##   left daughter right daughter split var split point status prediction
## 1             2              3        V7    0.037175      1       <NA>
## 2             4              5       V33   -0.099865      1       <NA>
## 3             6              7        V4   -0.609635      1       <NA>
## 4             8              9       V33   -0.777780      1       <NA>
## 5            10             11       V32   -0.118790      1       <NA>
## 6             0              0      <NA>    0.000000     -1          0

## Warning in RColorBrewer::brewer.pal(nlevs, "Set1"): minimal value for n is 3, returning requested pal
```


**OOB of randomforest**
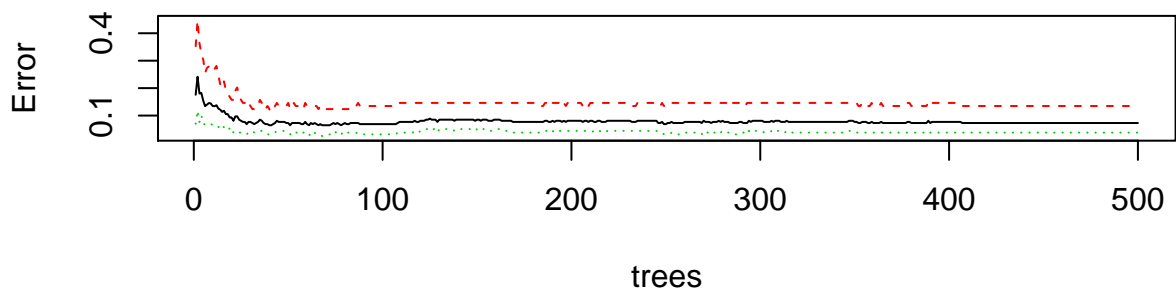
Boundary point detection
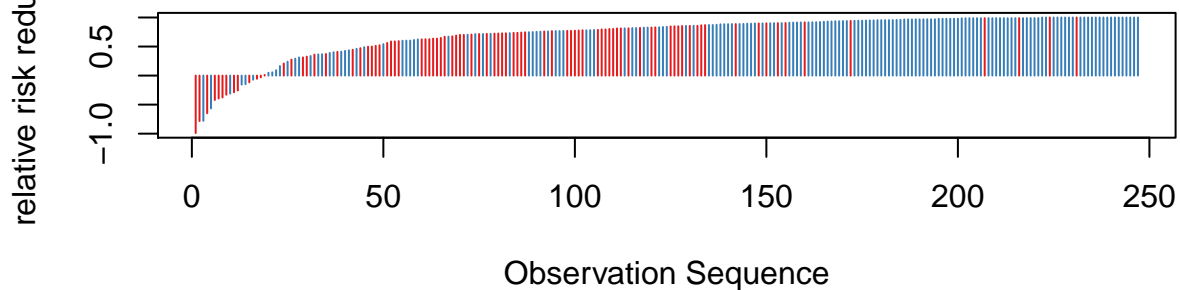
```
## Warning in par(DrawL): graphical parameter "cin" cannot be set
```

```
## Warning in par(DrawL): graphical parameter "cra" cannot be set

## Warning in par(DrawL): graphical parameter "csi" cannot be set

## Warning in par(DrawL): graphical parameter "cxy" cannot be set

## Warning in par(DrawL): graphical parameter "din" cannot be set

## Warning in par(DrawL): graphical parameter "page" cannot be set
```

Prediction accuracy change
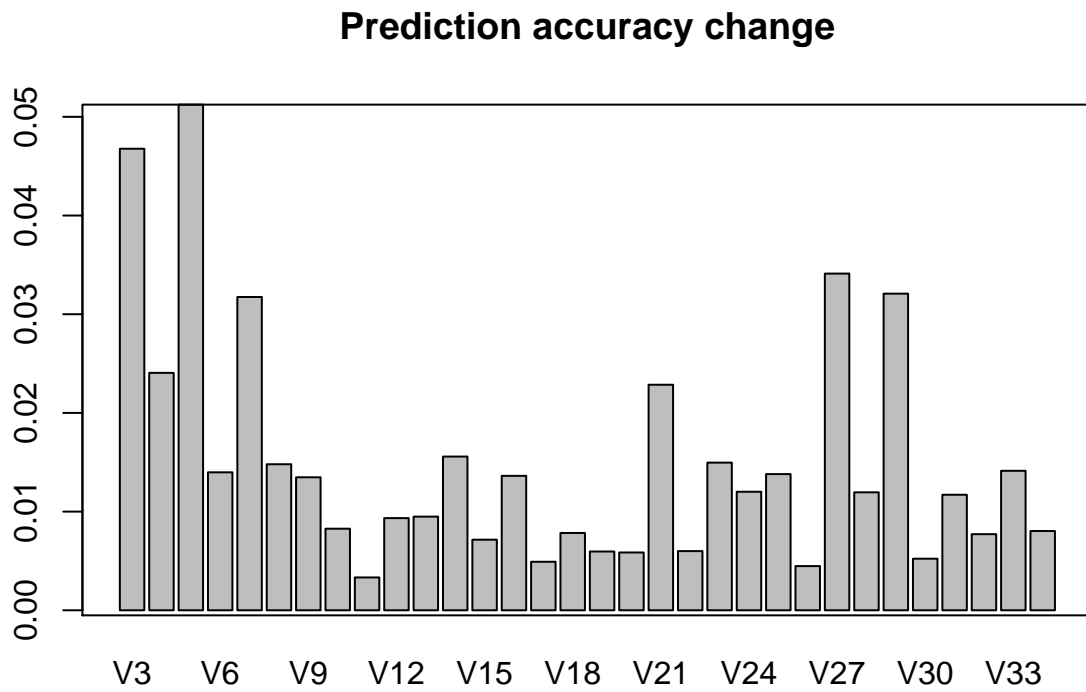
Input variable importance measure scatter plot
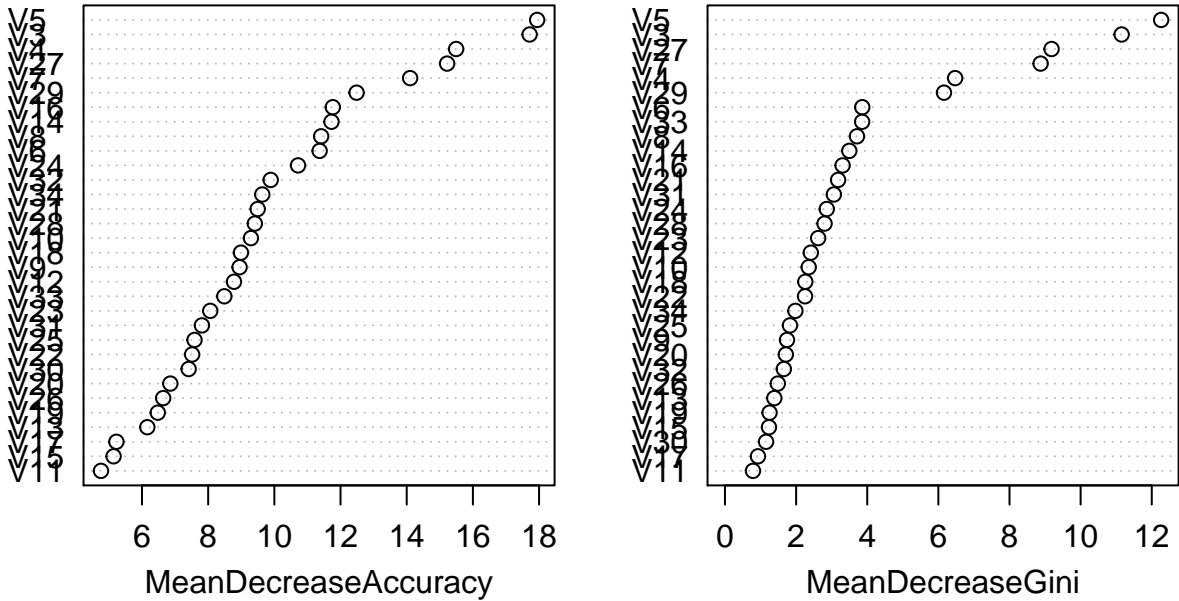

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 32  2
##          1  5 65
##
##                Accuracy : 0.9327
##                  95% CI : (0.8662, 0.9725)
##     No Information Rate : 0.6442
##     P-Value [Acc > NIR] : 5.263e-12
##
##                   Kappa : 0.8505
##
##  Mcnemar's Test P-Value : 0.4497
##
##             Sensitivity : 0.8649
##             Specificity : 0.9701
##          Pos Pred Value : 0.9412
##          Neg Pred Value : 0.9286
##              Prevalence : 0.3558
##          Detection Rate : 0.3077
##    Detection Prevalence : 0.3269
##       Balanced Accuracy : 0.9175
##
##        'Positive' Class : 0
```

##

Perform model selection and choose the best model for this data set.

```
## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases
```
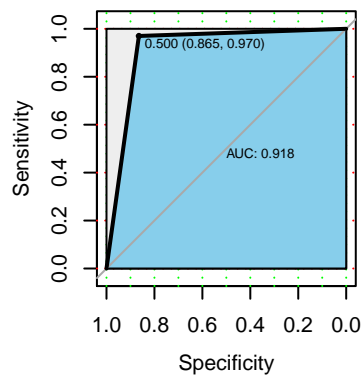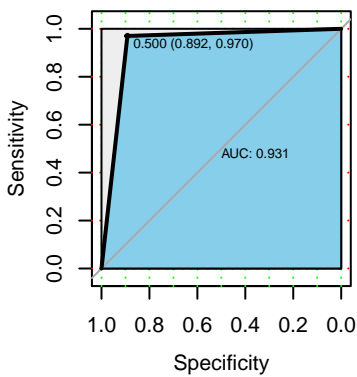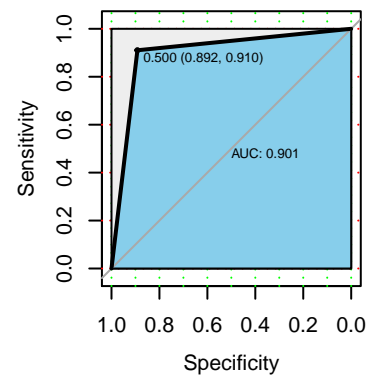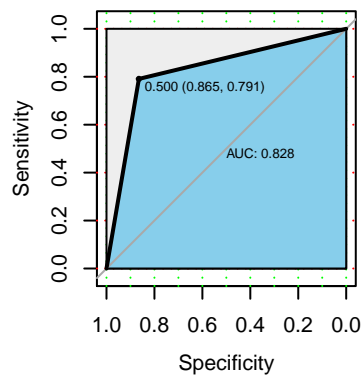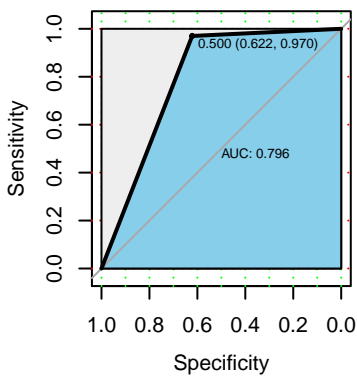
Based on the above result, the SVM model give the best prediction result, becuase the model has highest accuracy and the largest AUC value..

I am using the dataset about the Human resource investigation to identifiv the influence factors of a colleague leaving the company. I used the method of logistic regression and the CART decision tree. Because They give a clear result in classification scenario. According to the results, the CART decision tree gives a better result compared with the logistic forest.

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: left
##
## Terms added sequentially (first to last)
##
##
##                      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                                 1904     2167.9
## satisfaction_level    1   316.58      1903     1851.3 < 2.2e-16 ***
## last_evaluation       1     7.59      1902     1843.8 0.0058732 **
## number_project        1     2.76      1901     1841.0 0.0967678 .
## average_montly_hours  1    18.97      1900     1822.0 1.331e-05 ***
## time_spend_company    1    94.01      1899     1728.0 < 2.2e-16 ***
## Work_accident         1    49.24      1898     1678.8 2.265e-12 ***
## promotion_last_5years 1     2.36      1897     1676.4 0.1241328
## salary                2    17.73      1895     1658.7 0.0001416 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


##
## Call:
## glm(formula = left ~ ., family = binomial(link = "logit"), data = traindata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4904  -0.6497  -0.4037   0.6768   2.4570
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -2.442293   0.564521  -4.326 1.52e-05 ***
## satisfaction_level    -4.089367   0.267096 -15.310  < 2e-16 ***
## last_evaluation        0.501423   0.410832   1.221 0.222273
## number_project        -0.300292   0.058071  -5.171 2.33e-07 ***
## average_montly_hours   0.006447   0.001427   4.519 6.22e-06 ***
## time_spend_company     0.450450   0.050333   8.949  < 2e-16 ***
## Work_accident.L        1.080653   0.177405   6.091 1.12e-09 ***
## promotion_last_5years.L -0.871242  0.599203  -1.454 0.145946
## salary.L              -0.786852   0.212584  -3.701 0.000214 ***
## salary.Q              -0.474975   0.144506  -3.287 0.001013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 2167.9  on 1904  degrees of freedom
## Residual deviance: 1658.7  on 1895  degrees of freedom
## AIC: 1678.7
##
## Number of Fisher Scoring iterations: 5


## Start:  AIC=1678.69
## left ~ satisfaction_level + last_evaluation + number_project +
##     average_montly_hours + time_spend_company + Work_accident +
##     promotion_last_5years + salary
##
##                         Df Deviance    AIC
## - last_evaluation        1   1660.2 1678.2
## <none>                       1658.7 1678.7
## - promotion_last_5years  1   1661.3 1679.3
## - salary                 2   1676.4 1692.4
## - average_montly_hours   1   1679.6 1697.6
## - number_project         1   1686.7 1704.7
## - Work_accident          1   1707.1 1725.1
## - time_spend_company     1   1747.3 1765.3
## - satisfaction_level     1   1947.8 1965.8
##
## Step:  AIC=1678.18
## left ~ satisfaction_level + number_project + average_montly_hours +
##     time_spend_company + Work_accident + promotion_last_5years +
##     salary
##
##                         Df Deviance    AIC
## <none>                       1660.2 1678.2
## + last_evaluation        1   1658.7 1678.7
## - promotion_last_5years  1   1662.8 1678.8
## - salary                 2   1678.1 1692.1
## - number_project         1   1686.7 1702.7
## - average_montly_hours   1   1686.8 1702.8
## - Work_accident          1   1707.7 1723.7
## - time_spend_company     1   1752.8 1768.8
## - satisfaction_level     1   1951.2 1967.2


##
## Call:
## glm(formula = left ~ satisfaction_level + number_project + average_montly_hours +
##     time_spend_company + Work_accident + promotion_last_5years +
##     salary, family = binomial(link = "logit"), data = traindata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4944  -0.6460  -0.4026   0.6810   2.4727
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -2.303253   0.553287  -4.163 3.14e-05 ***
## satisfaction_level  -4.025094   0.260436 -15.455  < 2e-16 ***
## number_project      -0.282643   0.055983  -5.049 4.45e-07 ***
```

```
## average_montly_hours      0.006953   0.001368   5.083 3.72e-07 ***
## time_spend_company         0.456529   0.050109   9.111  < 2e-16 ***
## Work_accident.L            1.067600   0.177004   6.031 1.62e-09 ***
## promotion_last_5years.L   -0.869595   0.599804  -1.450 0.147115
## salary.L                  -0.791734   0.211948  -3.736 0.000187 ***
## salary.Q                  -0.471589   0.144192  -3.271 0.001073 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2167.9  on 1904  degrees of freedom
## Residual deviance: 1660.2  on 1896  degrees of freedom
## AIC: 1678.2
##
## Number of Fisher Scoring iterations: 5


## [1] 1678.183


## [1] 1678.692


## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0    1
##          0 557 147
##          1  49  62
##
##               Accuracy : 0.7595
##                 95% CI : (0.7286, 0.7885)
##    No Information Rate : 0.7436
##    P-Value [Acc > NIR] : 0.158
##
##                  Kappa : 0.255
##
##  Mcnemar's Test P-Value : 4.251e-12
##
##            Sensitivity : 0.9191
##            Specificity : 0.2967
##         Pos Pred Value : 0.7912
##         Neg Pred Value : 0.5586
##             Prevalence : 0.7436
##         Detection Rate : 0.6834
##   Detection Prevalence : 0.8638
##      Balanced Accuracy : 0.6079
##
##       'Positive' Class : 0
##


## [1] 75.95092


## Setting levels: control = 0, case = 1
```
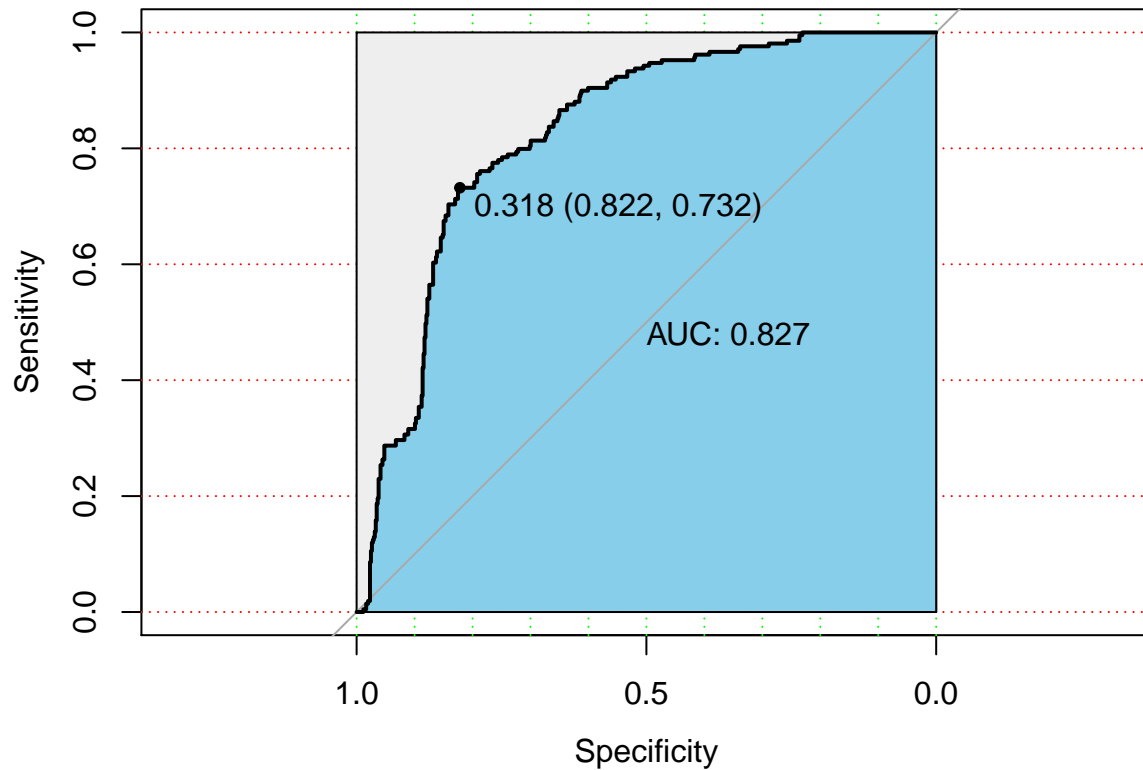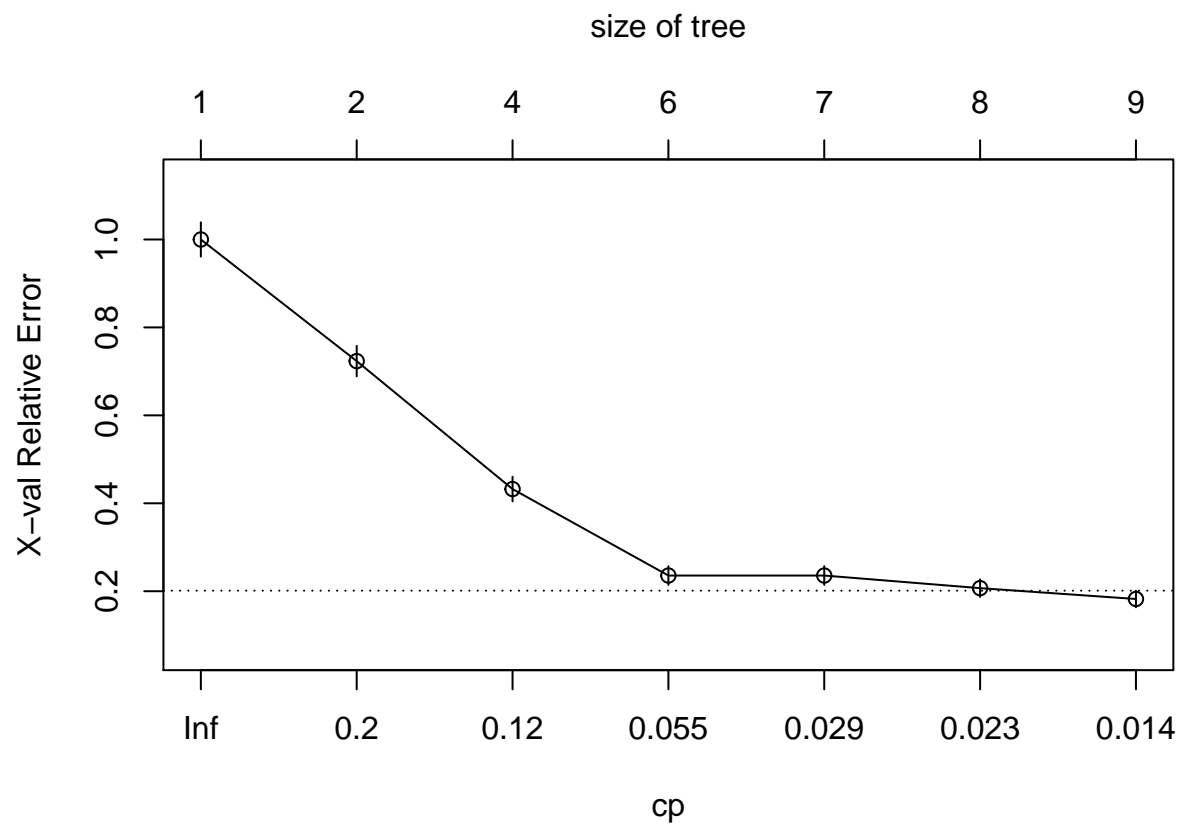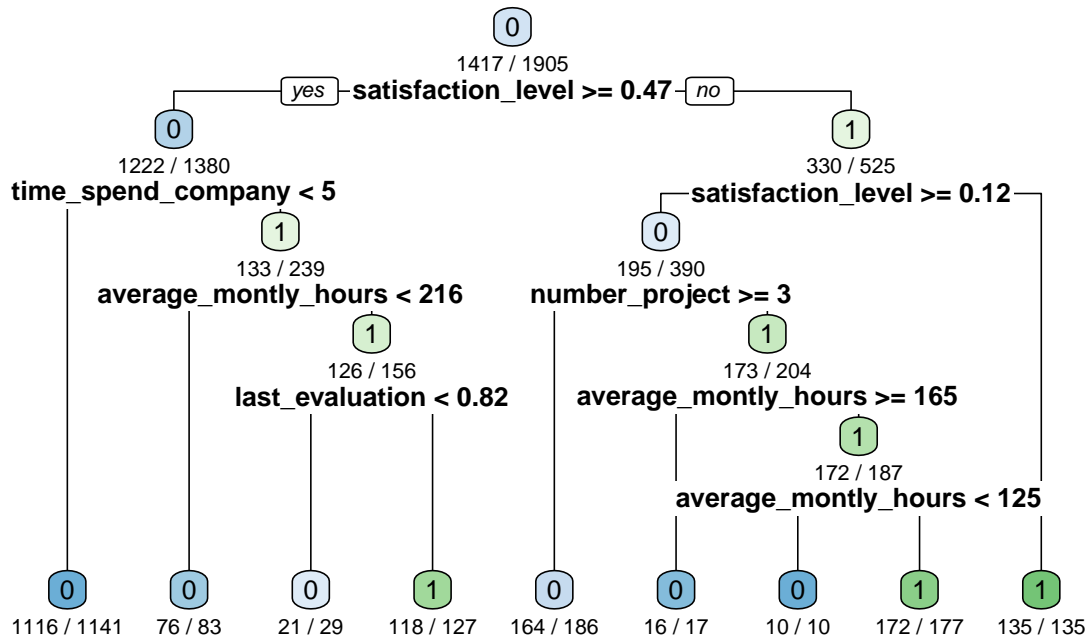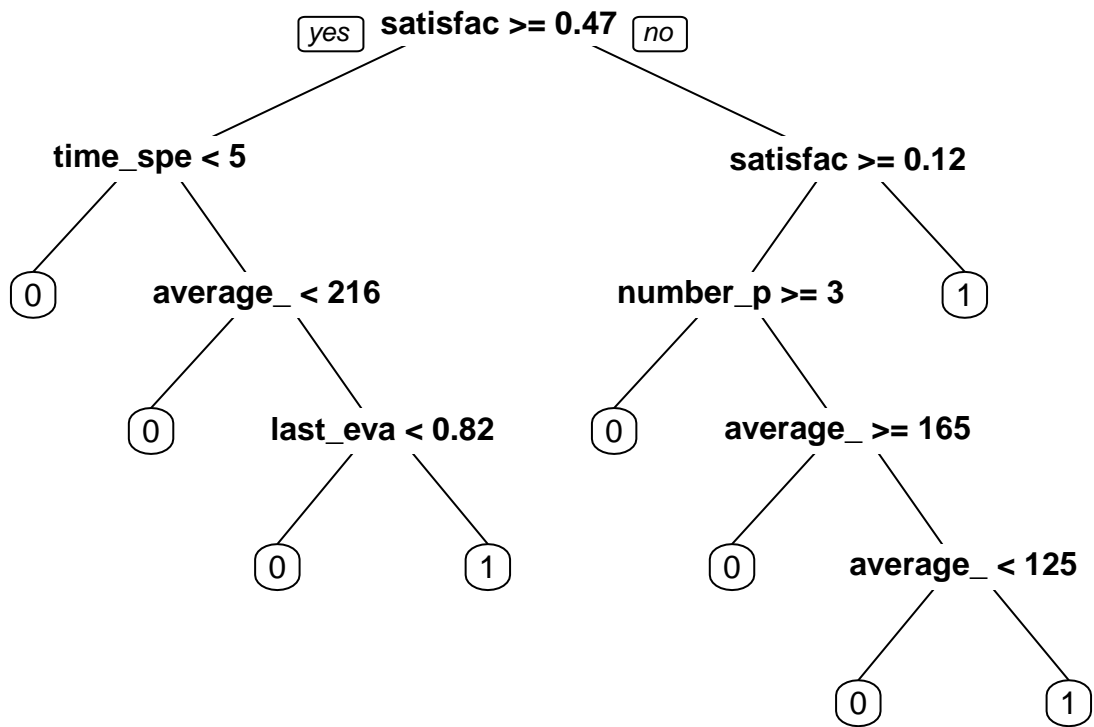
```
## 
## Classification tree:
## rpart(formula = left ~ ., data = traindata, method = "class", 
##     parms = list(split = "gini"))
## 
## Variables actually used in tree construction:
## [1] average_montly_hours last_evaluation       number_project
## [4] satisfaction_level   time_spend_company
## 
## Root node error: 488/1905 = 0.25617
## 
## n= 1905 
## 
##         CP nsplit rel error  xerror    xstd
## 1 0.276639      0   1.00000 1.00000 0.039042
## 2 0.145492      1   0.72336 0.72336 0.034751
## 3 0.098361      3   0.43238 0.43238 0.028069
## 4 0.030738      5   0.23566 0.23566 0.021301
## 5 0.026639      6   0.20492 0.23566 0.021301
## 6 0.020492      7   0.17828 0.20697 0.020041
## 7 0.010000      8   0.15779 0.18238 0.018875
```

Visualization of CART

0
1417 / 1905
yes — **satisfaction_level >= 0.47** — no

0
1222 / 1380
time_spend_company < 5

1
330 / 525
satisfaction_level >= 0.12

1
133 / 239
average_montly_hours < 216

0
195 / 390
number_project >= 3

1
126 / 156
last_evaluation < 0.82

1
173 / 204
average_montly_hours >= 165

1
172 / 187
average_montly_hours < 125

0
1116 / 1141

0
76 / 83

0
21 / 29

1
118 / 127

0
164 / 186

0
16 / 17

0
10 / 10

1
172 / 177

1
135 / 135

satisfac >= 0.47 yes no

time_spe < 5

satisfac >= 0.12

0

average_ < 216

number_p >= 3

1

0

last_eva < 0.82

0

average_ >= 165

0

1

0

average_ < 125

0

1

33

Visualization of CART
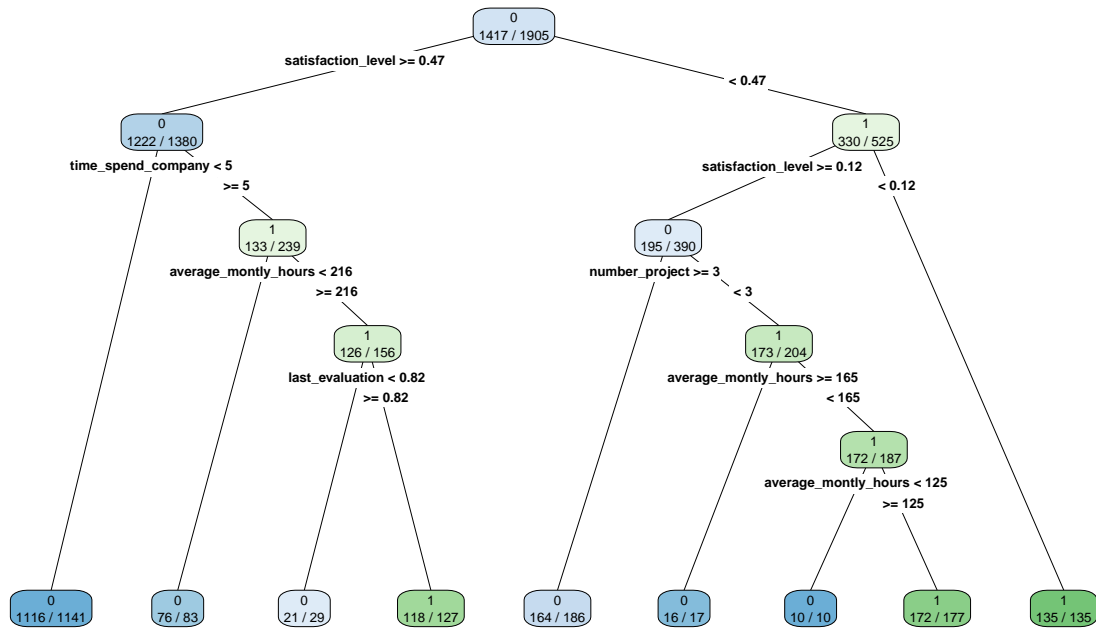


```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0    1
##          0 600   20
##          1   6  189
##
##               Accuracy : 0.9681
##                 95% CI : (0.9536, 0.9791)
##    No Information Rate : 0.7436
##    P-Value [Acc > NIR] : < 2e-16
##
##                  Kappa : 0.9145
##
##  Mcnemar's Test P-Value : 0.01079
##
##            Sensitivity : 0.9901
##            Specificity : 0.9043
##         Pos Pred Value : 0.9677
##         Neg Pred Value : 0.9692
##             Prevalence : 0.7436
##         Detection Rate : 0.7362
##   Detection Prevalence : 0.7607
##      Balanced Accuracy : 0.9472
##
##       'Positive' Class : 0
```

```
##

## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```