Final\_project\_all\_employees.R

Jennifer

Mon May 08 13:17:48 2017

library(caret)

## Warning: package 'caret' was built under R version 3.3.3

## Loading required package: lattice

## Loading required package: ggplot2

library(mlbench)

## Warning: package 'mlbench' was built under R version 3.3.3

library(randomForest)

## Warning: package 'randomForest' was built under R version 3.3.3

## randomForest 4.6-12

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

##   
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':  
##   
## margin

library(rpart)  
library(rpart.plot)

## Warning: package 'rpart.plot' was built under R version 3.3.3

library(party)

## Warning: package 'party' was built under R version 3.3.3

## Loading required package: grid

## Loading required package: mvtnorm

## Loading required package: modeltools

## Loading required package: stats4

## Loading required package: strucchange

## Warning: package 'strucchange' was built under R version 3.3.3

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

## Loading required package: sandwich

library(partykit)

## Warning: package 'partykit' was built under R version 3.3.3

##   
## Attaching package: 'partykit'

## The following objects are masked from 'package:party':  
##   
## cforest, ctree, ctree\_control, edge\_simple, mob, mob\_control,  
## node\_barplot, node\_bivplot, node\_boxplot, node\_inner,  
## node\_surv, node\_terminal

library(gmodels)

## Warning: package 'gmodels' was built under R version 3.3.3

library(e1071)

## Warning: package 'e1071' was built under R version 3.3.3

library(knitr)

## Warning: package 'knitr' was built under R version 3.3.3

options(scipen = 999)  
  
setwd("C:/Users/Jennifer/Documents/ADM/Final")  
HR<-read.csv("WatsonHR.csv")  
names(HR)

## [1] "Age" "Attrition"   
## [3] "BusinessTravel" "DailyRate"   
## [5] "Department" "DistanceFromHome"   
## [7] "Education" "EducationField"   
## [9] "EmployeeCount" "EmployeeNumber"   
## [11] "EnvironmentSatisfaction" "Gender"   
## [13] "HourlyRate" "JobInvolvement"   
## [15] "JobLevel" "JobRole"   
## [17] "JobSatisfaction" "MaritalStatus"   
## [19] "MonthlyIncome" "MonthlyRate"   
## [21] "NumCompaniesWorked" "Over18"   
## [23] "OverTime" "PercentSalaryHike"   
## [25] "PerformanceRating" "RelationshipSatisfaction"  
## [27] "StandardHours" "StockOptionLevel"   
## [29] "TotalWorkingYears" "TrainingTimesLastYear"   
## [31] "WorkLifeBalance" "YearsAtCompany"   
## [33] "YearsInCurrentRole" "YearsSinceLastPromotion"   
## [35] "YearsWithCurrManager"

HR<-HR[,c(2, 1, 3:8, 11:21, 23:26, 28:35)]  
summary(HR)

## Attrition Age BusinessTravel DailyRate   
## No :1233 Min. :18.00 Non-Travel : 150 Min. : 102.0   
## Yes: 237 1st Qu.:30.00 Travel\_Frequently: 277 1st Qu.: 465.0   
## Median :36.00 Travel\_Rarely :1043 Median : 802.0   
## Mean :36.92 Mean : 802.5   
## 3rd Qu.:43.00 3rd Qu.:1157.0   
## Max. :60.00 Max. :1499.0   
##   
## Department DistanceFromHome Education   
## Human Resources : 63 Min. : 1.000 Min. :1.000   
## Research & Development:961 1st Qu.: 2.000 1st Qu.:2.000   
## Sales :446 Median : 7.000 Median :3.000   
## Mean : 9.193 Mean :2.913   
## 3rd Qu.:14.000 3rd Qu.:4.000   
## Max. :29.000 Max. :5.000   
##   
## EducationField EnvironmentSatisfaction Gender   
## Human Resources : 27 Min. :1.000 Female:588   
## Life Sciences :606 1st Qu.:2.000 Male :882   
## Marketing :159 Median :3.000   
## Medical :464 Mean :2.722   
## Other : 82 3rd Qu.:4.000   
## Technical Degree:132 Max. :4.000   
##   
## HourlyRate JobInvolvement JobLevel   
## Min. : 30.00 Min. :1.00 Min. :1.000   
## 1st Qu.: 48.00 1st Qu.:2.00 1st Qu.:1.000   
## Median : 66.00 Median :3.00 Median :2.000   
## Mean : 65.89 Mean :2.73 Mean :2.064   
## 3rd Qu.: 83.75 3rd Qu.:3.00 3rd Qu.:3.000   
## Max. :100.00 Max. :4.00 Max. :5.000   
##   
## JobRole JobSatisfaction MaritalStatus  
## Sales Executive :326 Min. :1.000 Divorced:327   
## Research Scientist :292 1st Qu.:2.000 Married :673   
## Laboratory Technician :259 Median :3.000 Single :470   
## Manufacturing Director :145 Mean :2.729   
## Healthcare Representative:131 3rd Qu.:4.000   
## Manager :102 Max. :4.000   
## (Other) :215   
## MonthlyIncome MonthlyRate NumCompaniesWorked OverTime   
## Min. : 1009 Min. : 2094 Min. :0.000 No :1054   
## 1st Qu.: 2911 1st Qu.: 8047 1st Qu.:1.000 Yes: 416   
## Median : 4919 Median :14236 Median :2.000   
## Mean : 6503 Mean :14313 Mean :2.693   
## 3rd Qu.: 8379 3rd Qu.:20462 3rd Qu.:4.000   
## Max. :19999 Max. :26999 Max. :9.000   
##   
## PercentSalaryHike PerformanceRating RelationshipSatisfaction  
## Min. :11.00 Min. :3.000 Min. :1.000   
## 1st Qu.:12.00 1st Qu.:3.000 1st Qu.:2.000   
## Median :14.00 Median :3.000 Median :3.000   
## Mean :15.21 Mean :3.154 Mean :2.712   
## 3rd Qu.:18.00 3rd Qu.:3.000 3rd Qu.:4.000   
## Max. :25.00 Max. :4.000 Max. :4.000   
##   
## StockOptionLevel TotalWorkingYears TrainingTimesLastYear WorkLifeBalance  
## Min. :0.0000 Min. : 0.00 Min. :0.000 Min. :1.000   
## 1st Qu.:0.0000 1st Qu.: 6.00 1st Qu.:2.000 1st Qu.:2.000   
## Median :1.0000 Median :10.00 Median :3.000 Median :3.000   
## Mean :0.7939 Mean :11.28 Mean :2.799 Mean :2.761   
## 3rd Qu.:1.0000 3rd Qu.:15.00 3rd Qu.:3.000 3rd Qu.:3.000   
## Max. :3.0000 Max. :40.00 Max. :6.000 Max. :4.000   
##   
## YearsAtCompany YearsInCurrentRole YearsSinceLastPromotion  
## Min. : 0.000 Min. : 0.000 Min. : 0.000   
## 1st Qu.: 3.000 1st Qu.: 2.000 1st Qu.: 0.000   
## Median : 5.000 Median : 3.000 Median : 1.000   
## Mean : 7.008 Mean : 4.229 Mean : 2.188   
## 3rd Qu.: 9.000 3rd Qu.: 7.000 3rd Qu.: 3.000   
## Max. :40.000 Max. :18.000 Max. :15.000   
##   
## YearsWithCurrManager  
## Min. : 0.000   
## 1st Qu.: 2.000   
## Median : 3.000   
## Mean : 4.123   
## 3rd Qu.: 7.000   
## Max. :17.000   
##

dim(HR)

## [1] 1470 31

str(HR)

## 'data.frame': 1470 obs. of 31 variables:  
## $ Attrition : Factor w/ 2 levels "No","Yes": 2 1 2 1 1 1 1 1 1 1 ...  
## $ Age : int 41 49 37 33 27 32 59 30 38 36 ...  
## $ BusinessTravel : Factor w/ 3 levels "Non-Travel","Travel\_Frequently",..: 3 2 3 2 3 2 3 3 2 3 ...  
## $ DailyRate : int 1102 279 1373 1392 591 1005 1324 1358 216 1299 ...  
## $ Department : Factor w/ 3 levels "Human Resources",..: 3 2 2 2 2 2 2 2 2 2 ...  
## $ DistanceFromHome : int 1 8 2 3 2 2 3 24 23 27 ...  
## $ Education : int 2 1 2 4 1 2 3 1 3 3 ...  
## $ EducationField : Factor w/ 6 levels "Human Resources",..: 2 2 5 2 4 2 4 2 2 4 ...  
## $ EnvironmentSatisfaction : int 2 3 4 4 1 4 3 4 4 3 ...  
## $ Gender : Factor w/ 2 levels "Female","Male": 1 2 2 1 2 2 1 2 2 2 ...  
## $ HourlyRate : int 94 61 92 56 40 79 81 67 44 94 ...  
## $ JobInvolvement : int 3 2 2 3 3 3 4 3 2 3 ...  
## $ JobLevel : int 2 2 1 1 1 1 1 1 3 2 ...  
## $ JobRole : Factor w/ 9 levels "Healthcare Representative",..: 8 7 3 7 3 3 3 3 5 1 ...  
## $ JobSatisfaction : int 4 2 3 3 2 4 1 3 3 3 ...  
## $ MaritalStatus : Factor w/ 3 levels "Divorced","Married",..: 3 2 3 2 2 3 2 1 3 2 ...  
## $ MonthlyIncome : int 5993 5130 2090 2909 3468 3068 2670 2693 9526 5237 ...  
## $ MonthlyRate : int 19479 24907 2396 23159 16632 11864 9964 13335 8787 16577 ...  
## $ NumCompaniesWorked : int 8 1 6 1 9 0 4 1 0 6 ...  
## $ OverTime : Factor w/ 2 levels "No","Yes": 2 1 2 2 1 1 2 1 1 1 ...  
## $ PercentSalaryHike : int 11 23 15 11 12 13 20 22 21 13 ...  
## $ PerformanceRating : int 3 4 3 3 3 3 4 4 4 3 ...  
## $ RelationshipSatisfaction: int 1 4 2 3 4 3 1 2 2 2 ...  
## $ StockOptionLevel : int 0 1 0 0 1 0 3 1 0 2 ...  
## $ TotalWorkingYears : int 8 10 7 8 6 8 12 1 10 17 ...  
## $ TrainingTimesLastYear : int 0 3 3 3 3 2 3 2 2 3 ...  
## $ WorkLifeBalance : int 1 3 3 3 3 2 2 3 3 2 ...  
## $ YearsAtCompany : int 6 10 0 8 2 7 1 1 9 7 ...  
## $ YearsInCurrentRole : int 4 7 0 7 2 7 0 0 7 7 ...  
## $ YearsSinceLastPromotion : int 0 1 0 3 2 3 0 0 1 7 ...  
## $ YearsWithCurrManager : int 5 7 0 0 2 6 0 0 8 7 ...

names(HR)

## [1] "Attrition" "Age"   
## [3] "BusinessTravel" "DailyRate"   
## [5] "Department" "DistanceFromHome"   
## [7] "Education" "EducationField"   
## [9] "EnvironmentSatisfaction" "Gender"   
## [11] "HourlyRate" "JobInvolvement"   
## [13] "JobLevel" "JobRole"   
## [15] "JobSatisfaction" "MaritalStatus"   
## [17] "MonthlyIncome" "MonthlyRate"   
## [19] "NumCompaniesWorked" "OverTime"   
## [21] "PercentSalaryHike" "PerformanceRating"   
## [23] "RelationshipSatisfaction" "StockOptionLevel"   
## [25] "TotalWorkingYears" "TrainingTimesLastYear"   
## [27] "WorkLifeBalance" "YearsAtCompany"   
## [29] "YearsInCurrentRole" "YearsSinceLastPromotion"   
## [31] "YearsWithCurrManager"

set.seed(123)  
HR\_rand <- HR[order(runif(1470)), ]  
1470\*.8

## [1] 1176

HR\_train <- HR\_rand[1:1176,]  
HR\_test <- HR\_rand[1177:1470,]  
  
prop.table(table(HR\_train$Attrition))

##   
## No Yes   
## 0.8384354 0.1615646

prop.table(table(HR\_test$Attrition))

##   
## No Yes   
## 0.8401361 0.1598639

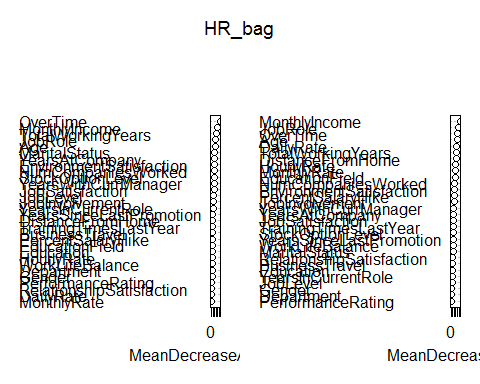
##################################################################  
#Random Forest  
  
set.seed(123)   
HR\_bag <- randomForest(Attrition~., data=HR\_train, mtry=29, na.action=na.omit, importance=TRUE)  
print(HR\_bag)

##   
## Call:  
## randomForest(formula = Attrition ~ ., data = HR\_train, mtry = 29, importance = TRUE, na.action = na.omit)   
## Type of random forest: classification  
## Number of trees: 500  
## No. of variables tried at each split: 29  
##   
## OOB estimate of error rate: 13.86%  
## Confusion matrix:  
## No Yes class.error  
## No 966 20 0.02028398  
## Yes 143 47 0.75263158

importance(HR\_bag)

## No Yes MeanDecreaseAccuracy  
## Age 9.16554575 7.90522548 12.2691569  
## BusinessTravel 1.72482917 4.04900604 3.2360481  
## DailyRate -1.99843137 0.48055142 -1.4488248  
## Department 0.34046597 1.11190950 0.9436414  
## DistanceFromHome 1.70637422 3.59627103 3.4852634  
## Education 3.38630243 -1.40197693 2.5195039  
## EducationField 3.25076047 0.31312731 2.9873957  
## EnvironmentSatisfaction 4.21743421 8.79231716 8.1009092  
## Gender 0.76907619 0.05000603 0.6602940  
## HourlyRate 2.33240237 0.45188233 2.1587434  
## JobInvolvement 3.67538840 6.01289479 6.0313395  
## JobLevel 3.92102563 7.42115023 7.1424125  
## JobRole 9.31655025 9.91904250 13.2123199  
## JobSatisfaction 5.39363942 5.46734103 7.3223546  
## MaritalStatus 6.75059170 7.12318627 9.0485512  
## MonthlyIncome 11.02763344 14.84559566 17.9107938  
## MonthlyRate -2.33898448 -0.86051661 -2.3541921  
## NumCompaniesWorked 7.47781208 2.49736419 8.0584005  
## OverTime 19.23545414 25.45156523 28.8954514  
## PercentSalaryHike 3.65375367 -0.21429480 3.1496278  
## PerformanceRating -0.43910808 1.75996875 0.3922708  
## RelationshipSatisfaction 0.05854356 0.63958186 0.3114183  
## StockOptionLevel 6.05417318 6.67091775 7.9344438  
## TotalWorkingYears 11.86288962 5.50520800 14.3017879  
## TrainingTimesLastYear 2.52173393 2.17257094 3.2360779  
## WorkLifeBalance 0.32979442 3.97916233 2.0383513  
## YearsAtCompany 7.80117387 4.01877273 8.7199317  
## YearsInCurrentRole 4.40217279 3.02508780 5.9375835  
## YearsSinceLastPromotion 3.65427633 1.22807520 3.7864679  
## YearsWithCurrManager 6.37483018 3.43851237 7.5200782  
## MeanDecreaseGini  
## Age 19.3311665  
## BusinessTravel 5.4108164  
## DailyRate 16.8623653  
## Department 1.0090959  
## DistanceFromHome 15.9022293  
## Education 4.8984059  
## EducationField 12.5022739  
## EnvironmentSatisfaction 11.6452966  
## Gender 1.3488513  
## HourlyRate 14.9340314  
## JobInvolvement 9.1286064  
## JobLevel 4.1450433  
## JobRole 21.7283593  
## JobSatisfaction 8.7107081  
## MaritalStatus 6.8622544  
## MonthlyIncome 29.1571714  
## MonthlyRate 13.8299103  
## NumCompaniesWorked 11.9310974  
## OverTime 21.0168168  
## PercentSalaryHike 9.9271161  
## PerformanceRating 0.5978799  
## RelationshipSatisfaction 5.5798185  
## StockOptionLevel 7.9588834  
## TotalWorkingYears 16.7542346  
## TrainingTimesLastYear 8.5774528  
## WorkLifeBalance 7.4470857  
## YearsAtCompany 8.8598578  
## YearsInCurrentRole 4.4965149  
## YearsSinceLastPromotion 7.8628336  
## YearsWithCurrManager 8.8614998

varImpPlot(HR\_bag)



actualRF <- HR\_test$Attrition   
predictedRF <- predict(HR\_bag, HR\_test, type="class")   
HR\_bag\_matrix <- confusionMatrix(predictedRF, actualRF, positive="Yes")   
print("Bagged results")

## [1] "Bagged results"

print(HR\_bag\_matrix)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 244 36  
## Yes 3 11  
##   
## Accuracy : 0.8673   
## 95% CI : (0.8231, 0.9039)  
## No Information Rate : 0.8401   
## P-Value [Acc > NIR] : 0.1146   
##   
## Kappa : 0.31   
## Mcnemar's Test P-Value : 0.000000299   
##   
## Sensitivity : 0.23404   
## Specificity : 0.98785   
## Pos Pred Value : 0.78571   
## Neg Pred Value : 0.87143   
## Prevalence : 0.15986   
## Detection Rate : 0.03741   
## Detection Prevalence : 0.04762   
## Balanced Accuracy : 0.61095   
##   
## 'Positive' Class : Yes   
##

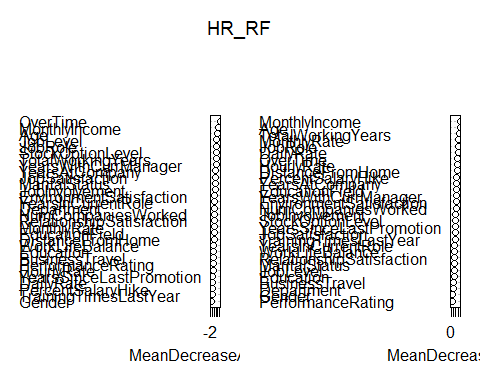
HR\_RF <- randomForest(Attrition~., data=HR\_train, mtry=3, ntree=100, na.action=na.omit, importance=TRUE)  
print(HR\_RF)

##   
## Call:  
## randomForest(formula = Attrition ~ ., data = HR\_train, mtry = 3, ntree = 100, importance = TRUE, na.action = na.omit)   
## Type of random forest: classification  
## Number of trees: 100  
## No. of variables tried at each split: 3  
##   
## OOB estimate of error rate: 14.54%  
## Confusion matrix:  
## No Yes class.error  
## No 978 8 0.00811359  
## Yes 163 27 0.85789474

importance(HR\_RF)

## No Yes MeanDecreaseAccuracy  
## Age 3.76386632 3.30366061 5.31870045  
## BusinessTravel 1.19512282 0.32544811 1.05213969  
## DailyRate -0.19082447 0.40148706 0.05464839  
## Department 1.45719225 1.86457024 2.11174254  
## DistanceFromHome 1.90409563 0.48698512 1.76443554  
## Education 1.18845708 0.28794988 1.26898702  
## EducationField 1.42933177 1.27488421 1.84359669  
## EnvironmentSatisfaction 1.11886482 3.09003118 2.43331330  
## Gender -1.52663027 -0.74945239 -1.76789466  
## HourlyRate 0.48820102 0.69739495 0.75641949  
## JobInvolvement 2.30627277 2.72102574 3.10475263  
## JobLevel 3.20298214 4.16460281 4.96791226  
## JobRole 4.27066098 2.90237602 4.51302077  
## JobSatisfaction 2.62397220 3.01466852 3.45284010  
## MaritalStatus 2.30983046 3.11369627 3.27523691  
## MonthlyIncome 4.24961099 4.72818930 6.41645118  
## MonthlyRate 1.66457171 0.93260335 1.85592089  
## NumCompaniesWorked 2.10553383 0.41290944 2.09623229  
## OverTime 6.84844284 7.06381800 9.17720875  
## PercentSalaryHike -0.30106224 0.74734733 -0.04980674  
## PerformanceRating -0.05533035 1.88484077 0.90964000  
## RelationshipSatisfaction 2.31898976 -0.77103588 1.92701959  
## StockOptionLevel 2.94566951 3.69369122 4.16667552  
## TotalWorkingYears 2.60140830 3.89603863 4.06391004  
## TrainingTimesLastYear -0.63383442 0.05104966 -0.47888238  
## WorkLifeBalance 1.13794120 1.39200668 1.46248500  
## YearsAtCompany 2.51854571 1.82389301 3.55317600  
## YearsInCurrentRole 0.62580049 3.72779081 2.32533768  
## YearsSinceLastPromotion 0.73022047 -1.23839605 0.11640543  
## YearsWithCurrManager 3.10264230 2.28870321 3.94434255  
## MeanDecreaseGini  
## Age 19.511664  
## BusinessTravel 5.478954  
## DailyRate 15.475292  
## Department 4.102394  
## DistanceFromHome 13.450792  
## Education 6.728688  
## EducationField 11.136995  
## EnvironmentSatisfaction 10.752798  
## Gender 2.595337  
## HourlyRate 13.608471  
## JobInvolvement 9.001969  
## JobLevel 6.961548  
## JobRole 15.748762  
## JobSatisfaction 8.709488  
## MaritalStatus 7.068804  
## MonthlyIncome 20.261001  
## MonthlyRate 16.467346  
## NumCompaniesWorked 10.580237  
## OverTime 14.132955  
## PercentSalaryHike 12.447346  
## PerformanceRating 2.475534  
## RelationshipSatisfaction 7.209054  
## StockOptionLevel 8.758900  
## TotalWorkingYears 17.641425  
## TrainingTimesLastYear 8.486839  
## WorkLifeBalance 7.784110  
## YearsAtCompany 11.519841  
## YearsInCurrentRole 8.260761  
## YearsSinceLastPromotion 8.724944  
## YearsWithCurrManager 11.062241

varImpPlot(HR\_RF)



actualRF2 <- HR\_test$Attrition   
predictedRF2 <- predict(HR\_RF, HR\_test, type="class")   
HR\_RF\_matrix <- confusionMatrix(predictedRF2, actualRF2, positive="Yes")   
print("RF results")

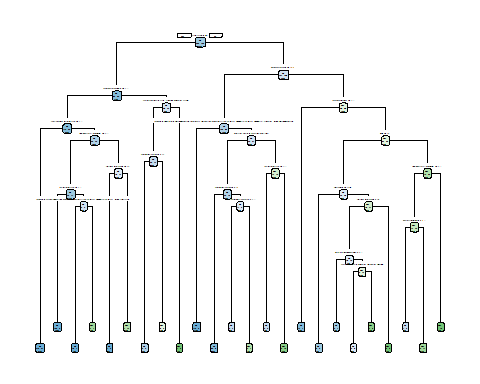
## [1] "RF results"

print(HR\_RF\_matrix)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 247 41  
## Yes 0 6  
##   
## Accuracy : 0.8605   
## 95% CI : (0.8156, 0.898)  
## No Information Rate : 0.8401   
## P-Value [Acc > NIR] : 0.1918   
##   
## Kappa : 0.1974   
## Mcnemar's Test P-Value : 0.0000000004185  
##   
## Sensitivity : 0.12766   
## Specificity : 1.00000   
## Pos Pred Value : 1.00000   
## Neg Pred Value : 0.85764   
## Prevalence : 0.15986   
## Detection Rate : 0.02041   
## Detection Prevalence : 0.02041   
## Balanced Accuracy : 0.56383   
##   
## 'Positive' Class : Yes   
##

##################################################################  
#Decision Tree  
  
set.seed(123)  
HR\_DT <- rpart(HR\_train$Attrition~., method="class", parms = list(split="gini"), data=HR\_train)  
rpart.plot(HR\_DT, type=1, extra=101)

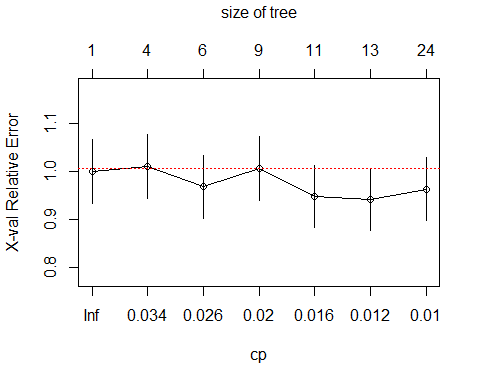
## Warning: labs do not fit even at cex 0.15, there may be some overplotting



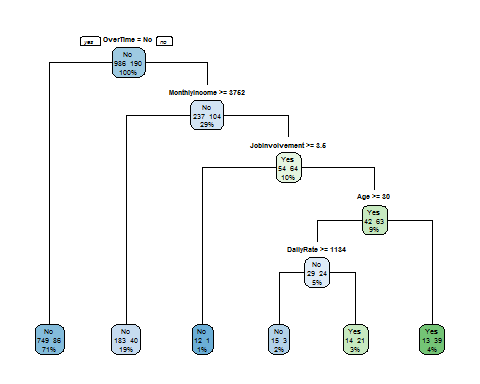
set.seed(123)  
cptable<-printcp(HR\_DT)

##   
## Classification tree:  
## rpart(formula = HR\_train$Attrition ~ ., data = HR\_train, method = "class",   
## parms = list(split = "gini"))  
##   
## Variables actually used in tree construction:  
## [1] Age DailyRate   
## [3] DistanceFromHome EducationField   
## [5] EnvironmentSatisfaction JobInvolvement   
## [7] JobRole JobSatisfaction   
## [9] MaritalStatus MonthlyIncome   
## [11] NumCompaniesWorked OverTime   
## [13] PercentSalaryHike TotalWorkingYears   
## [15] WorkLifeBalance YearsSinceLastPromotion  
## [17] YearsWithCurrManager   
##   
## Root node error: 190/1176 = 0.16156  
##   
## n= 1176   
##   
## CP nsplit rel error xerror xstd  
## 1 0.036842 0 1.00000 1.00000 0.066429  
## 2 0.031579 3 0.88947 1.01053 0.066710  
## 3 0.021053 5 0.82632 0.96842 0.065570  
## 4 0.018421 8 0.76316 1.00526 0.066570  
## 5 0.013158 10 0.72632 0.94737 0.064984  
## 6 0.010526 12 0.70000 0.94211 0.064836  
## 7 0.010000 23 0.58421 0.96316 0.065425

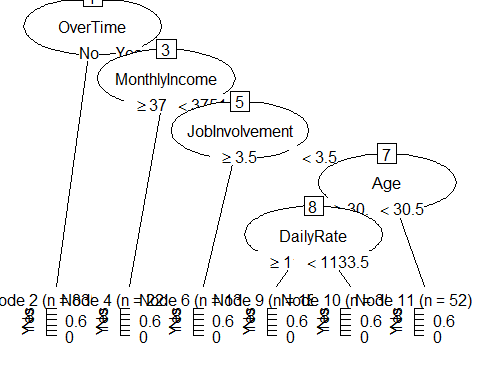
plotcp(HR\_DT, minline=TRUE, col="red")



#set.seed(123)  
#Pruned\_HR\_DT <-prune(HR\_DT,cp=.032, minsplit=10, minbucket=round(minsplit/3))   
#rpart.plot(Pruned\_HR\_DT, type=1, extra=101)  
#Pruned\_HR\_DT\_party<-as.party(Pruned\_HR\_DT)  
#plot(Pruned\_HR\_DT\_party)  
  
set.seed(123)  
Pruned\_HR\_DT2 <-prune(HR\_DT,cp=.022, minsplit=10, minbucket=round(minsplit/3))   
rpart.plot(Pruned\_HR\_DT2, type=1, extra=101)



Pruned\_HR\_DT2\_party2<-as.party(Pruned\_HR\_DT2)  
plot(Pruned\_HR\_DT2\_party2)



summary(Pruned\_HR\_DT2)

## Call:  
## rpart(formula = HR\_train$Attrition ~ ., data = HR\_train, method = "class",   
## parms = list(split = "gini"))  
## n= 1176   
##   
## CP nsplit rel error xerror xstd  
## 1 0.03684211 0 1.0000000 1.0000000 0.06642904  
## 2 0.03157895 3 0.8894737 1.0105263 0.06670999  
## 3 0.02200000 5 0.8263158 0.9684211 0.06557035  
##   
## Variable importance  
## MonthlyIncome OverTime JobLevel   
## 21 18 16   
## JobRole Age TotalWorkingYears   
## 12 9 9   
## JobInvolvement DailyRate YearsAtCompany   
## 6 4 3   
## NumCompaniesWorked Education MonthlyRate   
## 1 1 1   
##   
## Node number 1: 1176 observations, complexity param=0.03684211  
## predicted class=No expected loss=0.1615646 P(node) =1  
## class counts: 986 190  
## probabilities: 0.838 0.162   
## left son=2 (835 obs) right son=3 (341 obs)  
## Primary splits:  
## OverTime splits as LR, improve=19.75736, (0 missing)  
## TotalWorkingYears < 1.5 to the right, improve=17.71487, (0 missing)  
## JobLevel < 1.5 to the right, improve=15.62217, (0 missing)  
## YearsWithCurrManager < 0.5 to the right, improve=15.42283, (0 missing)  
## MonthlyIncome < 2439.5 to the right, improve=14.99433, (0 missing)  
## Surrogate splits:  
## DailyRate < 105 to the right, agree=0.711, adj=0.003, (0 split)  
##   
## Node number 2: 835 observations  
## predicted class=No expected loss=0.102994 P(node) =0.710034  
## class counts: 749 86  
## probabilities: 0.897 0.103   
##   
## Node number 3: 341 observations, complexity param=0.03684211  
## predicted class=No expected loss=0.3049853 P(node) =0.289966  
## class counts: 237 104  
## probabilities: 0.695 0.305   
## left son=6 (223 obs) right son=7 (118 obs)  
## Primary splits:  
## MonthlyIncome < 3751.5 to the right, improve=20.33655, (0 missing)  
## JobLevel < 1.5 to the right, improve=19.55813, (0 missing)  
## JobRole splits as LRRLLLRRR, improve=14.32438, (0 missing)  
## TotalWorkingYears < 8.5 to the right, improve=13.59519, (0 missing)  
## MaritalStatus splits as LLR, improve=11.89425, (0 missing)  
## Surrogate splits:  
## JobLevel < 1.5 to the right, agree=0.947, adj=0.847, (0 split)  
## JobRole splits as LRRLLLRLR, agree=0.862, adj=0.602, (0 split)  
## TotalWorkingYears < 8.5 to the right, agree=0.789, adj=0.390, (0 split)  
## Age < 28.5 to the right, agree=0.739, adj=0.246, (0 split)  
## YearsAtCompany < 1.5 to the right, agree=0.704, adj=0.144, (0 split)  
##   
## Node number 6: 223 observations  
## predicted class=No expected loss=0.1793722 P(node) =0.1896259  
## class counts: 183 40  
## probabilities: 0.821 0.179   
##   
## Node number 7: 118 observations, complexity param=0.03684211  
## predicted class=Yes expected loss=0.4576271 P(node) =0.1003401  
## class counts: 54 64  
## probabilities: 0.458 0.542   
## left son=14 (13 obs) right son=15 (105 obs)  
## Primary splits:  
## JobInvolvement < 3.5 to the right, improve=6.330117, (0 missing)  
## Age < 33.5 to the right, improve=5.732224, (0 missing)  
## YearsWithCurrManager < 0.5 to the right, improve=4.717135, (0 missing)  
## EnvironmentSatisfaction < 1.5 to the right, improve=4.591414, (0 missing)  
## MonthlyIncome < 2475 to the right, improve=4.545680, (0 missing)  
##   
## Node number 14: 13 observations  
## predicted class=No expected loss=0.07692308 P(node) =0.01105442  
## class counts: 12 1  
## probabilities: 0.923 0.077   
##   
## Node number 15: 105 observations, complexity param=0.03157895  
## predicted class=Yes expected loss=0.4 P(node) =0.08928571  
## class counts: 42 63  
## probabilities: 0.400 0.600   
## left son=30 (53 obs) right son=31 (52 obs)  
## Primary splits:  
## Age < 30.5 to the right, improve=4.635849, (0 missing)  
## EnvironmentSatisfaction < 1.5 to the right, improve=4.300281, (0 missing)  
## YearsWithCurrManager < 0.5 to the right, improve=4.200000, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve=4.187731, (0 missing)  
## NumCompaniesWorked < 0.5 to the left, improve=3.795722, (0 missing)  
## Surrogate splits:  
## TotalWorkingYears < 2.5 to the right, agree=0.657, adj=0.308, (0 split)  
## MonthlyIncome < 2422 to the right, agree=0.629, adj=0.250, (0 split)  
## NumCompaniesWorked < 1.5 to the right, agree=0.629, adj=0.250, (0 split)  
## Education < 3.5 to the right, agree=0.590, adj=0.173, (0 split)  
## MonthlyRate < 21874.5 to the left, agree=0.590, adj=0.173, (0 split)  
##   
## Node number 30: 53 observations, complexity param=0.03157895  
## predicted class=No expected loss=0.4528302 P(node) =0.04506803  
## class counts: 29 24  
## probabilities: 0.547 0.453   
## left son=60 (18 obs) right son=61 (35 obs)  
## Primary splits:  
## DailyRate < 1133.5 to the right, improve=4.464151, (0 missing)  
## PercentSalaryHike < 21.5 to the left, improve=3.358595, (0 missing)  
## PerformanceRating < 3.5 to the left, improve=1.975689, (0 missing)  
## JobRole splits as -LR---L-L, improve=1.913274, (0 missing)  
## MonthlyRate < 4623.5 to the right, improve=1.664151, (0 missing)  
## Surrogate splits:  
## MonthlyIncome < 2306.5 to the left, agree=0.736, adj=0.222, (0 split)  
## Age < 51.5 to the right, agree=0.698, adj=0.111, (0 split)  
## Department splits as RRL, agree=0.698, adj=0.111, (0 split)  
## JobRole splits as -RR---R-L, agree=0.698, adj=0.111, (0 split)  
## YearsAtCompany < 10.5 to the right, agree=0.698, adj=0.111, (0 split)  
##   
## Node number 31: 52 observations  
## predicted class=Yes expected loss=0.25 P(node) =0.04421769  
## class counts: 13 39  
## probabilities: 0.250 0.750   
##   
## Node number 60: 18 observations  
## predicted class=No expected loss=0.1666667 P(node) =0.01530612  
## class counts: 15 3  
## probabilities: 0.833 0.167   
##   
## Node number 61: 35 observations  
## predicted class=Yes expected loss=0.4 P(node) =0.0297619  
## class counts: 14 21  
## probabilities: 0.400 0.600

actualFullDT <- HR\_test$Attrition  
predictedFullDT <- predict(HR\_DT, HR\_test, type="class")  
results.matrix <- confusionMatrix(predictedFullDT, actualFullDT, positive="Yes")  
print(results.matrix)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 226 37  
## Yes 21 10  
##   
## Accuracy : 0.8027   
## 95% CI : (0.7526, 0.8467)  
## No Information Rate : 0.8401   
## P-Value [Acc > NIR] : 0.96349   
##   
## Kappa : 0.1482   
## Mcnemar's Test P-Value : 0.04888   
##   
## Sensitivity : 0.21277   
## Specificity : 0.91498   
## Pos Pred Value : 0.32258   
## Neg Pred Value : 0.85932   
## Prevalence : 0.15986   
## Detection Rate : 0.03401   
## Detection Prevalence : 0.10544   
## Balanced Accuracy : 0.56387   
##   
## 'Positive' Class : Yes   
##

#actualDT <- HR\_test$Attrition  
#predictedDT <- predict(Pruned\_HR\_DT, HR\_test, type="class")  
#DT.matrix <- confusionMatrix(predictedDT, actualDT, positive="Yes")  
#print(DT.matrix)  
  
actualDT2 <- HR\_test$Attrition  
predictedDT2 <- predict(Pruned\_HR\_DT2, HR\_test, type="class")  
DT.matrix2 <- confusionMatrix(predictedDT2, actualDT2, positive="Yes")  
print(DT.matrix2)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 239 36  
## Yes 8 11  
##   
## Accuracy : 0.8503   
## 95% CI : (0.8043, 0.8891)  
## No Information Rate : 0.8401   
## P-Value [Acc > NIR] : 0.3511   
##   
## Kappa : 0.2658   
## Mcnemar's Test P-Value : 0.00004693   
##   
## Sensitivity : 0.23404   
## Specificity : 0.96761   
## Pos Pred Value : 0.57895   
## Neg Pred Value : 0.86909   
## Prevalence : 0.15986   
## Detection Rate : 0.03741   
## Detection Prevalence : 0.06463   
## Balanced Accuracy : 0.60083   
##   
## 'Positive' Class : Yes   
##

#########################################################################################  
#Logistic Regression  
  
set.seed(123)  
HR\_train\_logit<-HR\_train  
names(HR\_train\_logit)

## [1] "Attrition" "Age"   
## [3] "BusinessTravel" "DailyRate"   
## [5] "Department" "DistanceFromHome"   
## [7] "Education" "EducationField"   
## [9] "EnvironmentSatisfaction" "Gender"   
## [11] "HourlyRate" "JobInvolvement"   
## [13] "JobLevel" "JobRole"   
## [15] "JobSatisfaction" "MaritalStatus"   
## [17] "MonthlyIncome" "MonthlyRate"   
## [19] "NumCompaniesWorked" "OverTime"   
## [21] "PercentSalaryHike" "PerformanceRating"   
## [23] "RelationshipSatisfaction" "StockOptionLevel"   
## [25] "TotalWorkingYears" "TrainingTimesLastYear"   
## [27] "WorkLifeBalance" "YearsAtCompany"   
## [29] "YearsInCurrentRole" "YearsSinceLastPromotion"   
## [31] "YearsWithCurrManager"

HR.logit <- glm(HR\_train\_logit$Attrition~., data=HR\_train\_logit, family=binomial())  
summary(HR.logit)

##   
## Call:  
## glm(formula = HR\_train\_logit$Attrition ~ ., family = binomial(),   
## data = HR\_train\_logit)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.8610 -0.4474 -0.2105 -0.0623 3.5223   
##   
## Coefficients:  
## Estimate Std. Error z value  
## (Intercept) -9.686089134 680.240307106 -0.014  
## Age -0.042585972 0.015921670 -2.675  
## BusinessTravelTravel\_Frequently 2.079128540 0.475855969 4.369  
## BusinessTravelTravel\_Rarely 1.065413405 0.431761618 2.468  
## DailyRate -0.000257236 0.000259050 -0.993  
## DepartmentResearch & Development 13.899428748 680.238563896 0.020  
## DepartmentSales 12.887578498 680.238974694 0.019  
## DistanceFromHome 0.044337116 0.012811409 3.461  
## Education -0.021621492 0.102859239 -0.210  
## EducationFieldLife Sciences -0.909631420 0.955153779 -0.952  
## EducationFieldMarketing -0.663201630 1.008885355 -0.657  
## EducationFieldMedical -0.866289690 0.952249746 -0.910  
## EducationFieldOther -1.018478701 1.011934875 -1.006  
## EducationFieldTechnical Degree 0.020666814 0.970534944 0.021  
## EnvironmentSatisfaction -0.555864220 0.098964110 -5.617  
## GenderMale 0.236262320 0.213621395 1.106  
## HourlyRate 0.004721530 0.005226240 0.903  
## JobInvolvement -0.681325913 0.145545313 -4.681  
## JobLevel -0.184278126 0.359898377 -0.512  
## JobRoleHuman Resources 15.111731381 680.238696129 0.022  
## JobRoleLaboratory Technician 1.636404259 0.580435929 2.819  
## JobRoleManager 1.076639526 0.979017718 1.100  
## JobRoleManufacturing Director -0.014845667 0.664073728 -0.022  
## JobRoleResearch Director -1.519501591 1.279220002 -1.188  
## JobRoleResearch Scientist 0.506475417 0.592924979 0.854  
## JobRoleSales Executive 2.291085228 1.418955000 1.615  
## JobRoleSales Representative 3.015658507 1.487242415 2.028  
## JobSatisfaction -0.498070216 0.095649825 -5.207  
## MaritalStatusMarried 0.175506451 0.311875565 0.563  
## MaritalStatusSingle 1.069541549 0.402051068 2.660  
## MonthlyIncome 0.000047178 0.000094247 0.501  
## MonthlyRate -0.000003462 0.000014645 -0.236  
## NumCompaniesWorked 0.191623071 0.046511467 4.120  
## OverTimeYes 2.111669365 0.229763709 9.191  
## PercentSalaryHike -0.038573106 0.046165576 -0.836  
## PerformanceRating 0.321192569 0.465034834 0.691  
## RelationshipSatisfaction -0.304165705 0.096778548 -3.143  
## StockOptionLevel -0.289244818 0.181731361 -1.592  
## TotalWorkingYears -0.071930034 0.033986029 -2.116  
## TrainingTimesLastYear -0.244690263 0.084211724 -2.906  
## WorkLifeBalance -0.595787733 0.146238830 -4.074  
## YearsAtCompany 0.026818026 0.049684490 0.540  
## YearsInCurrentRole -0.089954205 0.054414320 -1.653  
## YearsSinceLastPromotion 0.262002376 0.053847629 4.866  
## YearsWithCurrManager -0.150166082 0.056805318 -2.644  
## Pr(>|z|)   
## (Intercept) 0.988639   
## Age 0.007479 \*\*   
## BusinessTravelTravel\_Frequently 0.0000124680 \*\*\*  
## BusinessTravelTravel\_Rarely 0.013602 \*   
## DailyRate 0.320712   
## DepartmentResearch & Development 0.983698   
## DepartmentSales 0.984884   
## DistanceFromHome 0.000539 \*\*\*  
## Education 0.833508   
## EducationFieldLife Sciences 0.340924   
## EducationFieldMarketing 0.510949   
## EducationFieldMedical 0.362965   
## EducationFieldOther 0.314191   
## EducationFieldTechnical Degree 0.983011   
## EnvironmentSatisfaction 0.0000000194 \*\*\*  
## GenderMale 0.268732   
## HourlyRate 0.366299   
## JobInvolvement 0.0000028521 \*\*\*  
## JobLevel 0.608631   
## JobRoleHuman Resources 0.982276   
## JobRoleLaboratory Technician 0.004813 \*\*   
## JobRoleManager 0.271457   
## JobRoleManufacturing Director 0.982164   
## JobRoleResearch Director 0.234899   
## JobRoleResearch Scientist 0.392995   
## JobRoleSales Executive 0.106391   
## JobRoleSales Representative 0.042592 \*   
## JobSatisfaction 0.0000001917 \*\*\*  
## MaritalStatusMarried 0.573608   
## MaritalStatusSingle 0.007809 \*\*   
## MonthlyIncome 0.616672   
## MonthlyRate 0.813151   
## NumCompaniesWorked 0.0000379020 \*\*\*  
## OverTimeYes < 0.0000000000000002 \*\*\*  
## PercentSalaryHike 0.403415   
## PerformanceRating 0.489764   
## RelationshipSatisfaction 0.001673 \*\*   
## StockOptionLevel 0.111473   
## TotalWorkingYears 0.034306 \*   
## TrainingTimesLastYear 0.003665 \*\*   
## WorkLifeBalance 0.0000461979 \*\*\*  
## YearsAtCompany 0.589358   
## YearsInCurrentRole 0.098303 .   
## YearsSinceLastPromotion 0.0000011410 \*\*\*  
## YearsWithCurrManager 0.008205 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1040.18 on 1175 degrees of freedom  
## Residual deviance: 636.82 on 1131 degrees of freedom  
## AIC: 726.82  
##   
## Number of Fisher Scoring iterations: 15

#HR\_train\_logit2<-HR\_train\_logit[,c(1, 2, 12, 17, 20)]  
#HR.logit2 <- glm(HR\_train\_logit2$Attrition~., data=HR\_train\_logit2, family=binomial())  
#summary(HR.logit2)   
  
#HR\_train\_logit3<-HR\_train\_logit[,c(1, 2, 14, 17, 20, 25)]  
#HR.logit3 <- glm(HR\_train\_logit3$Attrition~., data=HR\_train\_logit3, family=binomial())  
#summary(HR.logit3)   
  
odds<-exp(cbind(Odds\_Ratio=coef(HR.logit)))   
odds

## Odds\_Ratio  
## (Intercept) 0.00006214196  
## Age 0.95830807415  
## BusinessTravelTravel\_Frequently 7.99749637770  
## BusinessTravelTravel\_Rarely 2.90203845323  
## DailyRate 0.99974279751  
## DepartmentResearch & Development 1087539.91841919604  
## DepartmentSales 395370.46803392371  
## DistanceFromHome 1.04533469459  
## Education 0.97861057654  
## EducationFieldLife Sciences 0.40267261373  
## EducationFieldMarketing 0.51519921387  
## EducationFieldMedical 0.42050887680  
## EducationFieldOther 0.36114393037  
## EducationFieldTechnical Degree 1.02088185179  
## EnvironmentSatisfaction 0.57357635054  
## GenderMale 1.26650649620  
## HourlyRate 1.00473269414  
## JobInvolvement 0.50594570720  
## JobLevel 0.83170445330  
## JobRoleHuman Resources 3655455.91723501496  
## JobRoleLaboratory Technician 5.13666614439  
## JobRoleManager 2.93480064058  
## JobRoleManufacturing Director 0.98526398694  
## JobRoleResearch Director 0.21882092199  
## JobRoleResearch Scientist 1.65943206928  
## JobRoleSales Executive 9.88566005911  
## JobRoleSales Representative 20.40252172032  
## JobSatisfaction 0.60770226322  
## MaritalStatusMarried 1.19184967707  
## MaritalStatusSingle 2.91404324762  
## MonthlyIncome 1.00004717865  
## MonthlyRate 0.99999653839  
## NumCompaniesWorked 1.21121388915  
## OverTimeYes 8.26202210602  
## PercentSalaryHike 0.96216136204  
## PerformanceRating 1.37877106344  
## RelationshipSatisfaction 0.73773860905  
## StockOptionLevel 0.74882885591  
## TotalWorkingYears 0.93059600359  
## TrainingTimesLastYear 0.78294700867  
## WorkLifeBalance 0.55112825311  
## YearsAtCompany 1.02718086510  
## YearsInCurrentRole 0.91397303992  
## YearsSinceLastPromotion 1.29952962991  
## YearsWithCurrManager 0.86056503988

prob<-odds/(1+odds)  
prob

## Odds\_Ratio  
## (Intercept) 0.0000621381  
## Age 0.4893551157  
## BusinessTravelTravel\_Frequently 0.8888579714  
## BusinessTravelTravel\_Rarely 0.7437236942  
## DailyRate 0.4999356911  
## DepartmentResearch & Development 0.9999990805  
## DepartmentSales 0.9999974707  
## DistanceFromHome 0.5110824636  
## Education 0.4945948375  
## EducationFieldLife Sciences 0.2870752660  
## EducationFieldMarketing 0.3400207769  
## EducationFieldMedical 0.2960269265  
## EducationFieldOther 0.2653238370  
## EducationFieldTechnical Degree 0.5051665197  
## EnvironmentSatisfaction 0.3645049383  
## GenderMale 0.5587923522  
## HourlyRate 0.5011803803  
## JobInvolvement 0.3359654367  
## JobLevel 0.4540603981  
## JobRoleHuman Resources 0.9999997264  
## JobRoleLaboratory Technician 0.8370450703  
## JobRoleManager 0.7458575182  
## JobRoleManufacturing Director 0.4962886515  
## JobRoleResearch Director 0.1795349243  
## JobRoleResearch Scientist 0.6239798671  
## JobRoleSales Executive 0.9081360253  
## JobRoleSales Representative 0.9532765338  
## JobSatisfaction 0.3779942823  
## MaritalStatusMarried 0.5437643327  
## MaritalStatusSingle 0.7445097213  
## MonthlyIncome 0.5000117944  
## MonthlyRate 0.4999991346  
## NumCompaniesWorked 0.5477597147  
## OverTimeYes 0.8920322162  
## PercentSalaryHike 0.4903579189  
## PerformanceRating 0.5796148627  
## RelationshipSatisfaction 0.4245394590  
## StockOptionLevel 0.4281887581  
## TotalWorkingYears 0.4820252408  
## TrainingTimesLastYear 0.4391308350  
## WorkLifeBalance 0.3553079844  
## YearsAtCompany 0.5067041046  
## YearsInCurrentRole 0.4775266009  
## YearsSinceLastPromotion 0.5651284563  
## YearsWithCurrManager 0.4625288670

anova(HR.logit,test="Chisq")

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: HR\_train\_logit$Attrition  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev  
## NULL 1175 1040.18  
## Age 1 41.254 1174 998.93  
## BusinessTravel 2 16.759 1172 982.17  
## DailyRate 1 3.074 1171 979.10  
## Department 2 7.933 1169 971.16  
## DistanceFromHome 1 6.738 1168 964.43  
## Education 1 0.068 1167 964.36  
## EducationField 5 7.807 1162 956.55  
## EnvironmentSatisfaction 1 21.418 1161 935.13  
## Gender 1 0.300 1160 934.83  
## HourlyRate 1 0.015 1159 934.82  
## JobInvolvement 1 28.034 1158 906.78  
## JobLevel 1 24.103 1157 882.68  
## JobRole 8 17.367 1149 865.31  
## JobSatisfaction 1 23.956 1148 841.36  
## MaritalStatus 2 23.130 1146 818.23  
## MonthlyIncome 1 1.555 1145 816.67  
## MonthlyRate 1 0.619 1144 816.06  
## NumCompaniesWorked 1 12.952 1143 803.10  
## OverTime 1 91.766 1142 711.34  
## PercentSalaryHike 1 0.271 1141 711.07  
## PerformanceRating 1 0.195 1140 710.87  
## RelationshipSatisfaction 1 7.152 1139 703.72  
## StockOptionLevel 1 3.971 1138 699.75  
## TotalWorkingYears 1 6.438 1137 693.31  
## TrainingTimesLastYear 1 8.675 1136 684.64  
## WorkLifeBalance 1 15.481 1135 669.15  
## YearsAtCompany 1 0.095 1134 669.06  
## YearsInCurrentRole 1 2.699 1133 666.36  
## YearsSinceLastPromotion 1 22.751 1132 643.61  
## YearsWithCurrManager 1 6.790 1131 636.82  
## Pr(>Chi)   
## NULL   
## Age 0.0000000001337 \*\*\*  
## BusinessTravel 0.0002295 \*\*\*  
## DailyRate 0.0795302 .   
## Department 0.0189404 \*   
## DistanceFromHome 0.0094378 \*\*   
## Education 0.7944723   
## EducationField 0.1671964   
## EnvironmentSatisfaction 0.0000036925346 \*\*\*  
## Gender 0.5841599   
## HourlyRate 0.9015385   
## JobInvolvement 0.0000001192304 \*\*\*  
## JobLevel 0.0000009129482 \*\*\*  
## JobRole 0.0265070 \*   
## JobSatisfaction 0.0000009857723 \*\*\*  
## MaritalStatus 0.0000094913354 \*\*\*  
## MonthlyIncome 0.2124533   
## MonthlyRate 0.4314713   
## NumCompaniesWorked 0.0003196 \*\*\*  
## OverTime < 0.00000000000000022 \*\*\*  
## PercentSalaryHike 0.6029329   
## PerformanceRating 0.6584384   
## RelationshipSatisfaction 0.0074862 \*\*   
## StockOptionLevel 0.0462923 \*   
## TotalWorkingYears 0.0111721 \*   
## TrainingTimesLastYear 0.0032259 \*\*   
## WorkLifeBalance 0.0000833308554 \*\*\*  
## YearsAtCompany 0.7583880   
## YearsInCurrentRole 0.1004094   
## YearsSinceLastPromotion 0.0000018437551 \*\*\*  
## YearsWithCurrManager 0.0091669 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

HR\_test\_logit<-HR\_test  
  
HR\_test\_logit$predict.Attrition<-predict(HR.logit, newdata=HR\_test\_logit,type = "response")  
  
HR\_test\_logit\_CI<-cbind(HR\_test\_logit,predict(HR.logit, newdata=HR\_test\_logit,type="link",se=TRUE))  
  
HR\_test\_logit\_CI <- within(HR\_test\_logit\_CI,   
 {  
 PredictedProb <- plogis(fit)  
 LL <- plogis(fit - (1.96 \* se.fit))  
 UL <- plogis(fit + (1.96 \* se.fit))  
 })   
  
names(HR\_test\_logit\_CI)

## [1] "Attrition" "Age"   
## [3] "BusinessTravel" "DailyRate"   
## [5] "Department" "DistanceFromHome"   
## [7] "Education" "EducationField"   
## [9] "EnvironmentSatisfaction" "Gender"   
## [11] "HourlyRate" "JobInvolvement"   
## [13] "JobLevel" "JobRole"   
## [15] "JobSatisfaction" "MaritalStatus"   
## [17] "MonthlyIncome" "MonthlyRate"   
## [19] "NumCompaniesWorked" "OverTime"   
## [21] "PercentSalaryHike" "PerformanceRating"   
## [23] "RelationshipSatisfaction" "StockOptionLevel"   
## [25] "TotalWorkingYears" "TrainingTimesLastYear"   
## [27] "WorkLifeBalance" "YearsAtCompany"   
## [29] "YearsInCurrentRole" "YearsSinceLastPromotion"   
## [31] "YearsWithCurrManager" "predict.Attrition"   
## [33] "fit" "se.fit"   
## [35] "residual.scale" "UL"   
## [37] "LL" "PredictedProb"

dim(HR\_test\_logit\_CI)

## [1] 294 38

HR\_test\_logit\_CI$predict.Attrition<-ifelse(HR\_test\_logit\_CI$predict.Attrition>.5, "yes", "no")  
probs<-HR\_test\_logit\_CI[,c(1,32)]   
  
CrossTable(x=probs$Attrition, y=probs$predict.Attrition, prob.chisq=FALSE)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 294   
##   
##   
## | probs$predict.Attrition   
## probs$Attrition | no | yes | Row Total |   
## ----------------|-----------|-----------|-----------|  
## No | 243 | 4 | 247 |   
## | 1.031 | 12.151 | |   
## | 0.984 | 0.016 | 0.840 |   
## | 0.897 | 0.174 | |   
## | 0.827 | 0.014 | |   
## ----------------|-----------|-----------|-----------|  
## Yes | 28 | 19 | 47 |   
## | 5.420 | 63.858 | |   
## | 0.596 | 0.404 | 0.160 |   
## | 0.103 | 0.826 | |   
## | 0.095 | 0.065 | |   
## ----------------|-----------|-----------|-----------|  
## Column Total | 271 | 23 | 294 |   
## | 0.922 | 0.078 | |   
## ----------------|-----------|-----------|-----------|  
##   
##

TP = 19  
TN = 243  
FP = 4  
FN = 28  
Sensitivity = TP/(TP+FN) #true positive rate; recall; TP/(TP+FN)  
Specificity = TN/(TN+FP) #how often is the prediction negative when actual is negative?  
  
Precision = TP/(TP+FP) #how often is prediction positive when actual is positive?  
Accuracy = (TP+TN)/(TP+TN+FP+FN) #how often is classifier correct  
Value<-round(c(TP,TN,FP,FN,Sensitivity,Specificity,Precision,Accuracy),digits=3)  
Measure<-c("True Positive","True Negative","False Positive","False Negative","Sensitivity=TP/(TN+FP)",  
 "Specificity=TN/(TN+TP)","Precision=TP/(TP+FP)","Accuracy=(TP+TN)/total")  
table<-as.data.frame(cbind(Measure,Value))  
kable(table)

|  |  |
| --- | --- |
| Measure | Value |
| True Positive | 19 |
| True Negative | 243 |
| False Positive | 4 |
| False Negative | 28 |
| Sensitivity=TP/(TN+FP) | 0.404 |
| Specificity=TN/(TN+TP) | 0.984 |
| Precision=TP/(TP+FP) | 0.826 |
| Accuracy=(TP+TN)/total | 0.891 |