Case 2: HR Analytics

Hunter Berberich

8/1/2024

Setup code

Add code here to load required packages, read data, etc. - preliminary tasks you need to perform to answer the HW question.

```
# Your code here.

# Load necessary libraries
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(caret)
```

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

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##      cov, smooth, var
```

```r
library(openxlsx)
library(rpart)
library(rpart.plot)
library(reshape2)
```

Data Cleaning and Preprocessing

```r
data <- read.csv("/Users/hunterberberich/Desktop/School Stuff!/Data/
    Case2DataClean.csv")

data$Education <- factor(data$Education, levels = 1:5, labels = c("Below
    College", "College", "Bachelor", "Master", "Doctor"))
data$EnvironmentSatisfaction <- factor(data$EnvironmentSatisfaction, levels =
    1:4, labels = c("Low", "Medium", "High", "Very High"))
data$JobInvolvement <- factor(data$JobInvolvement, levels = 1:4, labels = c("
    Low", "Medium", "High", "Very High"))
data$JobSatisfaction <- factor(data$JobSatisfaction, levels = 1:4, labels = c(
    "Low", "Medium", "High", "Very High"))
data$PerformanceRating <- factor(data$PerformanceRating, levels = 1:4, labels
    = c("Low", "Good", "Excellent", "Outstanding"))
data$RelationshipSatisfaction <- factor(data$RelationshipSatisfaction, levels
    = 1:4, labels = c("Low", "Medium", "High", "Very High"))
data$WorkLifeBalance <- factor(data$WorkLifeBalance, levels = 1:4, labels = c(
    "Bad", "Good", "Better", "Best"))
data$BusinessTravel <- factor(data$BusinessTravel)
data$Department <- factor(data$Department)
data$EducationField <- factor(data$EducationField)
data$Gender <- factor(data$Gender)
data$JobRole <- factor(data$JobRole)
data$MaritalStatus <- factor(data$MaritalStatus)
data$OverTime <- factor(data$OverTime)
data$Attrition <- factor(data$Attrition, levels = c("No", "Yes"))
```

Model Building

```r
set.seed(123)
index <- createDataPartition(data$Attrition, p = 0.8, list = FALSE)
train_data <- data[index, ]
test_data <- data[-index, ]

initial_model <- glm(Attrition ~ ., data = train_data, family = binomial)
print(initial_model)
```

```
##
## Call:  glm(formula = Attrition ~ ., family = binomial, data = train_data)
##
## Coefficients:
##                    (Intercept)                              Age
##                     -1.334e+01                       -2.609e-02
##   BusinessTravelTravel_Frequently        BusinessTravelTravel_Rarely
##                      2.346e+00                        1.384e+00
##                      DailyRate        DepartmentResearch & Development
##                     -3.693e-04                        1.339e+01
##                 DepartmentSales                   DistanceFromHome
##                      1.336e+01                        4.857e-02
##                 EducationCollege                   EducationBachelor
##                      6.716e-02                        1.600e-01
##                  EducationMaster                    EducationDoctor
##                     -2.156e-02                       -9.914e-01
##       EducationFieldLife Sciences           EducationFieldMarketing
##                     -8.673e-01                       -3.610e-01
##             EducationFieldMedical              EducationFieldOther
##                     -9.297e-01                       -9.623e-01
##     EducationFieldTechnical Degree      EnvironmentSatisfactionMedium
##                      1.620e-02                       -1.252e+00
##      EnvironmentSatisfactionHigh      EnvironmentSatisfactionVery High
##                     -1.279e+00                       -1.433e+00
##                     GenderMale                       HourlyRate
##                      5.360e-01                        1.693e-03
##            JobInvolvementMedium                JobInvolvementHigh
##                     -9.058e-01                       -1.225e+00
##          JobInvolvementVery High                       JobLevel
##                     -2.270e+00                        2.027e-01
##         JobRoleHuman Resources            JobRoleLaboratory Technician
##                      1.449e+01                        1.743e+00
##                 JobRoleManager            JobRoleManufacturing Director
##                     -3.285e-01                        5.859e-02
##          JobRoleResearch Director            JobRoleResearch Scientist
##                     -7.457e-01                        9.017e-01
##          JobRoleSales Executive            JobRoleSales Representative
##                      1.046e+00                        2.262e+00
##            JobSatisfactionMedium                JobSatisfactionHigh
##                     -8.207e-01                       -6.257e-01
##          JobSatisfactionVery High             MaritalStatusMarried
##                     -1.590e+00                        5.990e-01
##             MaritalStatusSingle                    MonthlyIncome
##                      1.679e+00                        3.579e-05
##                    MonthlyRate                  NumCompaniesWorked
##                      1.537e-06                        1.947e-01
##                     OverTimeYes                   PercentSalaryHike
##                      2.106e+00                       -2.053e-02
##      PerformanceRatingOutstanding       RelationshipSatisfactionMedium
##                      9.369e-02                       -7.462e-01
##     RelationshipSatisfactionHigh       RelationshipSatisfactionVery High
##                     -7.899e-01                       -8.433e-01
##                 StockOptionLevel                   TotalWorkingYears
##                      5.622e-02                       -1.095e-01
```

3

```
##                  TrainingTimesLastYear                    WorkLifeBalanceGood
##                            −8.334e−02                             −7.984e−01
##                  WorkLifeBalanceBetter                    WorkLifeBalanceBest
##                            −1.326e+00                             −9.584e−01
##                         YearsAtCompany                    YearsInCurrentRole
##                             1.133e−01                             −1.581e−01
##                 YearsSinceLastPromotion                   YearsWithCurrManager
##                             1.832e−01                             −1.398e−01
##
## Degrees of Freedom: 1176 Total (i.e. Null);   1119 Residual
## Null Deviance:          1041
## Residual Deviance: 654.4      AIC: 770.4
```

```
stepwise_model <- step(initial_model, direction = "both")
```

```
## Start:   AIC=770.44
## Attrition ~ Age + BusinessTravel + DailyRate + Department +
    DistanceFromHome +
##      Education + EducationField + EnvironmentSatisfaction + Gender +
##      HourlyRate + JobInvolvement + JobLevel + JobRole + JobSatisfaction +
##      MaritalStatus + MonthlyIncome + MonthlyRate + NumCompaniesWorked +
##      OverTime + PercentSalaryHike + PerformanceRating +
    RelationshipSatisfaction +
##      StockOptionLevel + TotalWorkingYears + TrainingTimesLastYear +
##      WorkLifeBalance + YearsAtCompany + YearsInCurrentRole +
    YearsSinceLastPromotion +
##      YearsWithCurrManager
##
##                            Df Deviance    AIC
## − Education                 4   657.44 765.44
## − Department                2   655.29 767.29
## − MonthlyRate               1   654.45 768.45
## − PerformanceRating         1   654.48 768.48
## − StockOptionLevel          1   654.54 768.54
## − HourlyRate                1   654.55 768.55
## − MonthlyIncome             1   654.59 768.59
## − PercentSalaryHike         1   654.65 768.65
## − JobLevel                  1   654.76 768.76
## − TrainingTimesLastYear     1   655.47 769.47
## − EducationField            5   663.47 769.47
## <none>                          654.44 770.44
## − DailyRate                 1   656.52 770.52
## − Age                       1   657.11 771.11
## − YearsAtCompany            1   660.04 774.04
## − RelationshipSatisfaction  3   664.36 774.36
## − Gender                    1   660.75 774.75
## − YearsWithCurrManager      1   661.14 775.14
## − JobRole                   8   675.28 775.28
## − YearsInCurrentRole        1   663.41 777.41
## − WorkLifeBalance           3   667.62 777.62
## − TotalWorkingYears         1   664.49 778.49
## − YearsSinceLastPromotion   1   668.65 782.65
## − JobInvolvement            3   672.93 782.93
## − DistanceFromHome          1   669.10 783.10
```

```
## - MaritalStatus              2   673.89  785.89
## - NumCompaniesWorked         1   672.74  786.74
## - EnvironmentSatisfaction    3   682.90  792.90
## - JobSatisfaction            3   684.02  794.02
## - BusinessTravel             2   682.36  794.36
## - OverTime                   1   749.97  863.97
##
## Step:   AIC=765.44
## Attrition ~ Age + BusinessTravel + DailyRate + Department +
    DistanceFromHome +
##      EducationField + EnvironmentSatisfaction + Gender + HourlyRate +
##      JobInvolvement + JobLevel + JobRole + JobSatisfaction + MaritalStatus +
##      MonthlyIncome + MonthlyRate + NumCompaniesWorked + OverTime +
##      PercentSalaryHike + PerformanceRating + RelationshipSatisfaction +
##      StockOptionLevel + TotalWorkingYears + TrainingTimesLastYear +
##      WorkLifeBalance + YearsAtCompany + YearsInCurrentRole +
    YearsSinceLastPromotion +
##      YearsWithCurrManager
##
##                             Df Deviance    AIC
## - Department                 2   658.23  762.23
## - MonthlyRate                1   657.44  763.44
## - PerformanceRating          1   657.49  763.49
## - HourlyRate                 1   657.52  763.52
## - StockOptionLevel           1   657.52  763.52
## - PercentSalaryHike          1   657.66  763.66
## - MonthlyIncome              1   657.69  763.69
## - JobLevel                   1   657.73  763.73
## - EducationField             5   666.47  764.47
## - TrainingTimesLastYear      1   658.52  764.52
## <none>                           657.44  765.44
## - DailyRate                  1   659.71  765.71
## - Age                        1   660.41  766.41
## - YearsAtCompany             1   662.46  768.46
## - Gender                     1   663.60  769.60
## - YearsWithCurrManager       1   663.79  769.79
## - RelationshipSatisfaction   3   667.87  769.87
## + Education                  4   654.44  770.44
## - JobRole                    8   679.29  771.29
## - YearsInCurrentRole         1   665.95  771.95
## - WorkLifeBalance            3   671.18  773.18
## - TotalWorkingYears          1   668.00  774.00
## - JobInvolvement             3   675.26  777.26
## - YearsSinceLastPromotion    1   671.61  777.61
## - DistanceFromHome           1   672.18  778.18
## - MaritalStatus              2   676.94  780.94
## - NumCompaniesWorked         1   675.81  781.81
## - EnvironmentSatisfaction    3   686.36  788.36
## - JobSatisfaction            3   687.18  789.18
## - BusinessTravel             2   685.60  789.60
## - OverTime                   1   751.13  857.13
##
## Step:   AIC=762.23
## Attrition ~ Age + BusinessTravel + DailyRate + DistanceFromHome +
```

```
##       EducationField + EnvironmentSatisfaction + Gender + HourlyRate +
##       JobInvolvement + JobLevel + JobRole + JobSatisfaction + MaritalStatus +
##       MonthlyIncome + MonthlyRate + NumCompaniesWorked + OverTime +
##       PercentSalaryHike + PerformanceRating + RelationshipSatisfaction +
##       StockOptionLevel + TotalWorkingYears + TrainingTimesLastYear +
##       WorkLifeBalance + YearsAtCompany + YearsInCurrentRole +
##    YearsSinceLastPromotion +
##       YearsWithCurrManager
##
##                              Df Deviance    AIC
## - MonthlyRate                 1   658.23 760.23
## - PerformanceRating           1   658.28 760.28
## - StockOptionLevel            1   658.29 760.29
## - HourlyRate                  1   658.32 760.32
## - PercentSalaryHike           1   658.43 760.43
## - MonthlyIncome               1   658.49 760.49
## - JobLevel                    1   658.51 760.51
## - EducationField              5   667.24 761.24
## - TrainingTimesLastYear       1   659.26 761.26
## <none>                            658.23 762.23
## - DailyRate                   1   660.55 762.55
## - Age                         1   661.22 763.22
## - YearsAtCompany              1   663.02 765.02
## + Department                  2   657.44 765.44
## - Gender                      1   664.34 766.34
## - YearsWithCurrManager        1   664.37 766.37
## - RelationshipSatisfaction    3   668.82 766.82
## + Education                   4   655.29 767.29
## - YearsInCurrentRole          1   666.52 768.52
## - WorkLifeBalance             3   672.12 770.12
## - TotalWorkingYears           1   668.84 770.84
## - JobRole                     8   684.36 772.36
## - JobInvolvement              3   676.31 774.31
## - YearsSinceLastPromotion     1   672.58 774.58
## - DistanceFromHome            1   673.02 775.02
## - MaritalStatus               2   677.60 777.60
## - NumCompaniesWorked          1   676.82 778.82
## - EnvironmentSatisfaction     3   687.25 785.25
## - JobSatisfaction             3   688.22 786.22
## - BusinessTravel              2   686.28 786.28
## - OverTime                    1   752.62 854.62
##
## Step:   AIC=760.23
## Attrition ~ Age + BusinessTravel + DailyRate + DistanceFromHome +
##       EducationField + EnvironmentSatisfaction + Gender + HourlyRate +
##       JobInvolvement + JobLevel + JobRole + JobSatisfaction + MaritalStatus +
##       MonthlyIncome + NumCompaniesWorked + OverTime + PercentSalaryHike +
##       PerformanceRating + RelationshipSatisfaction + StockOptionLevel +
##       TotalWorkingYears + TrainingTimesLastYear + WorkLifeBalance +
##       YearsAtCompany + YearsInCurrentRole + YearsSinceLastPromotion +
##       YearsWithCurrManager
##
##                              Df Deviance    AIC
## - PerformanceRating           1   658.28 758.28
```

6

```
## - StockOptionLevel              1     658.30  758.30
## - HourlyRate                    1     658.32  758.32
## - PercentSalaryHike             1     658.44  758.44
## - MonthlyIncome                 1     658.49  758.49
## - JobLevel                      1     658.52  758.52
## - EducationField                5     667.24  759.24
## - TrainingTimesLastYear         1     659.27  759.27
## <none>                                658.23  760.23
## - DailyRate                     1     660.56  760.56
## - Age                           1     661.23  761.23
## + MonthlyRate                   1     658.23  762.23
## - YearsAtCompany                1     663.02  763.02
## + Department                    2     657.44  763.44
## - Gender                        1     664.34  764.34
## - YearsWithCurrManager          1     664.39  764.39
## - RelationshipSatisfaction      3     668.87  764.87
## + Education                     4     655.30  765.30
## - YearsInCurrentRole            1     666.52  766.52
## - WorkLifeBalance               3     672.16  768.16
## - TotalWorkingYears             1     668.84  768.84
## - JobRole                       8     684.39  770.39
## - JobInvolvement                3     676.33  772.33
## - YearsSinceLastPromotion       1     672.63  772.63
## - DistanceFromHome              1     673.10  773.10
## - MaritalStatus                 2     677.61  775.61
## - NumCompaniesWorked            1     676.82  776.82
## - EnvironmentSatisfaction       3     687.25  783.25
## - JobSatisfaction               3     688.22  784.22
## - BusinessTravel                2     686.39  784.39
## - OverTime                      1     752.85  852.85
##
## Step:   AIC=758.28
## Attrition ~ Age + BusinessTravel + DailyRate + DistanceFromHome +
##     EducationField + EnvironmentSatisfaction + Gender + HourlyRate +
##     JobInvolvement + JobLevel + JobRole + JobSatisfaction + MaritalStatus +
##     MonthlyIncome + NumCompaniesWorked + OverTime + PercentSalaryHike +
##     RelationshipSatisfaction + StockOptionLevel + TotalWorkingYears +
##     TrainingTimesLastYear + WorkLifeBalance + YearsAtCompany +
##     YearsInCurrentRole + YearsSinceLastPromotion + YearsWithCurrManager
##
##                                Df Deviance     AIC
## - StockOptionLevel              1     658.34  756.34
## - HourlyRate                    1     658.36  756.36
## - PercentSalaryHike             1     658.48  756.48
## - MonthlyIncome                 1     658.52  756.52
## - JobLevel                      1     658.57  756.57
## - EducationField                5     667.28  757.28
## - TrainingTimesLastYear         1     659.31  757.31
## <none>                                658.28  758.28
## - DailyRate                     1     660.59  758.59
## - Age                           1     661.28  759.28
## + PerformanceRating             1     658.23  760.23
## + MonthlyRate                   1     658.28  760.28
## - YearsAtCompany                1     663.04  761.04
```

```
## + Department                     2    657.49  761.49
## - Gender                         1    664.36  762.36
## - YearsWithCurrManager           1    664.44  762.44
## - RelationshipSatisfaction       3    668.88  762.88
## + Education                      4    655.34  763.34
## - YearsInCurrentRole             1    666.53  764.53
## - WorkLifeBalance                3    672.26  766.26
## - TotalWorkingYears              1    668.84  766.84
## - JobRole                        8    684.44  768.44
## - JobInvolvement                 3    676.35  770.35
## - YearsSinceLastPromotion        1    672.74  770.74
## - DistanceFromHome               1    673.10  771.10
## - MaritalStatus                  2    677.62  773.62
## - NumCompaniesWorked             1    676.82  774.82
## - EnvironmentSatisfaction        3    687.34  781.34
## - JobSatisfaction                3    688.23  782.23
## - BusinessTravel                 2    686.52  782.52
## - OverTime                       1    753.05  851.05
##
## Step:   AIC=756.34
## Attrition ~ Age + BusinessTravel + DailyRate + DistanceFromHome +
##     EducationField + EnvironmentSatisfaction + Gender + HourlyRate +
##     JobInvolvement + JobLevel + JobRole + JobSatisfaction + MaritalStatus +
##     MonthlyIncome + NumCompaniesWorked + OverTime + PercentSalaryHike +
##     RelationshipSatisfaction + TotalWorkingYears + TrainingTimesLastYear +
##     WorkLifeBalance + YearsAtCompany + YearsInCurrentRole +
##  YearsSinceLastPromotion +
##     YearsWithCurrManager
##
##                               Df Deviance      AIC
## - HourlyRate                     1    658.43  754.43
## - PercentSalaryHike              1    658.53  754.53
## - MonthlyIncome                  1    658.57  754.57
## - JobLevel                       1    658.64  754.64
## - EducationField                 5    667.32  755.32
## - TrainingTimesLastYear          1    659.36  755.36
## <none>                                658.34  756.34
## - DailyRate                      1    660.64  756.64
## - Age                            1    661.32  757.32
## + StockOptionLevel               1    658.28  758.28
## + PerformanceRating              1    658.30  758.30
## + MonthlyRate                    1    658.34  758.34
## - YearsAtCompany                 1    663.09  759.09
## + Department                     2    657.57  759.57
## - YearsWithCurrManager           1    664.45  760.45
## - Gender                         1    664.47  760.47
## - RelationshipSatisfaction       3    669.03  761.03
## + Education                      4    655.41  761.41
## - YearsInCurrentRole             1    666.68  762.68
## - WorkLifeBalance                3    672.26  764.26
## - TotalWorkingYears              1    668.93  764.93
## - JobRole                        8    684.57  766.57
## - JobInvolvement                 3    676.54  768.54
## - YearsSinceLastPromotion        1    672.78  768.78
```

```
## − DistanceFromHome              1    673.29  769.29
## − NumCompaniesWorked            1    677.01  773.01
## − EnvironmentSatisfaction       3    687.34  779.34
## − JobSatisfaction               3    688.30  780.30
## − BusinessTravel                2    686.52  780.52
## − MaritalStatus                 2    692.95  786.95
## − OverTime                      1    753.21  849.21
##
## Step:   AIC=754.43
## Attrition ~ Age + BusinessTravel + DailyRate + DistanceFromHome +
##     EducationField + EnvironmentSatisfaction + Gender + JobInvolvement +
##     JobLevel + JobRole + JobSatisfaction + MaritalStatus + MonthlyIncome +
##     NumCompaniesWorked + OverTime + PercentSalaryHike +
##   RelationshipSatisfaction +
##     TotalWorkingYears + TrainingTimesLastYear + WorkLifeBalance +
##     YearsAtCompany + YearsInCurrentRole + YearsSinceLastPromotion +
##     YearsWithCurrManager
##
##                               Df Deviance    AIC
## − PercentSalaryHike            1    658.63  752.63
## − MonthlyIncome               1    658.66  752.66
## − JobLevel                    1    658.72  752.72
## − TrainingTimesLastYear       1    659.44  753.44
## − EducationField              5    667.44  753.44
## <none>                             658.43  754.43
## − DailyRate                   1    660.73  754.73
## − Age                         1    661.37  755.37
## + HourlyRate                  1    658.34  756.34
## + StockOptionLevel            1    658.36  756.36
## + PerformanceRating           1    658.39  756.39
## + MonthlyRate                 1    658.43  756.43
## − YearsAtCompany              1    663.25  757.25
## + Department                  2    657.66  757.66
## − Gender                      1    664.48  758.48
## − YearsWithCurrManager        1    664.53  758.53
## − RelationshipSatisfaction    3    669.07  759.07
## + Education                   4    655.52  759.52
## − YearsInCurrentRole          1    666.81  760.81
## − WorkLifeBalance             3    672.34  762.34
## − TotalWorkingYears           1    669.02  763.02
## − JobRole                     8    684.58  764.58
## − JobInvolvement              3    676.54  766.54
## − YearsSinceLastPromotion     1    672.78  766.78
## − DistanceFromHome            1    673.42  767.42
## − NumCompaniesWorked          1    677.08  771.08
## − EnvironmentSatisfaction     3    687.38  777.38
## − BusinessTravel              2    686.58  778.58
## − JobSatisfaction             3    688.77  778.77
## − MaritalStatus               2    693.00  785.00
## − OverTime                    1    753.29  847.29
##
## Step:   AIC=752.63
## Attrition ~ Age + BusinessTravel + DailyRate + DistanceFromHome +
##     EducationField + EnvironmentSatisfaction + Gender + JobInvolvement +
```

```
##          JobLevel + JobRole + JobSatisfaction + MaritalStatus + MonthlyIncome +
##          NumCompaniesWorked + OverTime + RelationshipSatisfaction +
##          TotalWorkingYears + TrainingTimesLastYear + WorkLifeBalance +
##          YearsAtCompany + YearsInCurrentRole + YearsSinceLastPromotion +
##          YearsWithCurrManager
##
##                                  Df  Deviance      AIC
## - MonthlyIncome                   1    658.85   750.85
## - JobLevel                        1    658.93   750.93
## - TrainingTimesLastYear           1    659.64   751.64
## - EducationField                  5    667.74   751.74
## <none>                                 658.63   752.63
## - DailyRate                       1    661.01   753.01
## - Age                             1    661.62   753.62
## + PercentSalaryHike               1    658.43   754.43
## + HourlyRate                      1    658.53   754.53
## + StockOptionLevel                1    658.57   754.57
## + PerformanceRating               1    658.58   754.58
## + MonthlyRate                     1    658.63   754.63
## - YearsAtCompany                  1    663.42   755.42
## + Department                      2    657.87   755.87
## - YearsWithCurrManager            1    664.72   756.72
## - Gender                          1    664.74   756.74
## - RelationshipSatisfaction        3    669.31   757.31
## + Education                       4    655.74   757.74
## - YearsInCurrentRole              1    667.03   759.03
## - WorkLifeBalance                 3    672.43   760.43
## - TotalWorkingYears               1    669.23   761.23
## - JobRole                         8    684.65   762.65
## - JobInvolvement                  3    676.70   764.70
## - YearsSinceLastPromotion         1    673.15   765.15
## - DistanceFromHome                1    673.52   765.52
## - NumCompaniesWorked              1    677.40   769.40
## - EnvironmentSatisfaction         3    687.53   775.53
## - JobSatisfaction                 3    689.02   777.02
## - BusinessTravel                  2    687.15   777.15
## - MaritalStatus                   2    693.29   783.29
## - OverTime                        1    753.31   845.31
##
## Step:   AIC=750.85
## Attrition ~ Age + BusinessTravel + DailyRate + DistanceFromHome +
##      EducationField + EnvironmentSatisfaction + Gender + JobInvolvement +
##      JobLevel + JobRole + JobSatisfaction + MaritalStatus +
##   NumCompaniesWorked +
##      OverTime + RelationshipSatisfaction + TotalWorkingYears +
##      TrainingTimesLastYear + WorkLifeBalance + YearsAtCompany +
##      YearsInCurrentRole + YearsSinceLastPromotion + YearsWithCurrManager
##
##                                  Df  Deviance      AIC
## - TrainingTimesLastYear           1    659.88   749.88
## - EducationField                  5    668.03   750.03
## - JobLevel                        1    660.17   750.17
## <none>                                 658.85   750.85
## - DailyRate                       1    661.25   751.25
```

```
## - Age                          1    661.86  751.86
## + MonthlyIncome                1    658.63  752.63
## + PercentSalaryHike            1    658.66  752.66
## + HourlyRate                   1    658.76  752.76
## + PerformanceRating            1    658.80  752.80
## + StockOptionLevel             1    658.80  752.80
## + MonthlyRate                  1    658.85  752.85
## - YearsAtCompany               1    663.64  753.64
## + Department                   2    658.08  754.08
## - Gender                       1    665.06  755.06
## - YearsWithCurrManager         1    665.18  755.18
## - RelationshipSatisfaction     3    669.52  755.52
## + Education                    4    655.86  755.86
## - YearsInCurrentRole           1    667.13  757.13
## - WorkLifeBalance              3    672.68  758.68
## - TotalWorkingYears            1    669.24  759.24
## - JobRole                      8    684.99  760.99
## - JobInvolvement               3    676.90  762.90
## - DistanceFromHome             1    673.54  763.54
## - YearsSinceLastPromotion      1    673.57  763.57
## - NumCompaniesWorked           1    677.89  767.89
## - EnvironmentSatisfaction      3    687.89  773.89
## - JobSatisfaction              3    689.27  775.27
## - BusinessTravel               2    687.33  775.33
## - MaritalStatus                2    693.41  781.41
## - OverTime                     1    753.87  843.87
##
## Step:   AIC=749.88
## Attrition ~ Age + BusinessTravel + DailyRate + DistanceFromHome +
##      EducationField + EnvironmentSatisfaction + Gender + JobInvolvement +
##      JobLevel + JobRole + JobSatisfaction + MaritalStatus +
##   NumCompaniesWorked +
##      OverTime + RelationshipSatisfaction + TotalWorkingYears +
##      WorkLifeBalance + YearsAtCompany + YearsInCurrentRole +
##   YearsSinceLastPromotion +
##      YearsWithCurrManager
##
##                          Df Deviance     AIC
## - EducationField          5    669.08  749.08
## - JobLevel                1    661.25  749.25
## <none>                         659.88  749.88
## - DailyRate               1    662.31  750.31
## + TrainingTimesLastYear   1    658.85  750.85
## - Age                     1    663.04  751.04
## + MonthlyIncome           1    659.64  751.64
## + PercentSalaryHike       1    659.70  751.70
## + HourlyRate              1    659.80  751.80
## + PerformanceRating       1    659.83  751.83
## + StockOptionLevel        1    659.84  751.84
## + MonthlyRate             1    659.88  751.88
## - YearsAtCompany          1    664.65  752.65
## + Department              2    659.17  753.17
## - YearsWithCurrManager    1    666.02  754.02
## - Gender                  1    666.27  754.27
```

```
## - RelationshipSatisfaction    3    670.72  754.72
## + Education                    4    656.83  754.83
## - YearsInCurrentRole           1    668.23  756.23
## - WorkLifeBalance              3    674.15  758.15
## - TotalWorkingYears            1    670.36  758.36
## - JobRole                      8    685.70  759.70
## - JobInvolvement               3    677.85  761.85
## - DistanceFromHome             1    674.48  762.48
## - YearsSinceLastPromotion      1    674.71  762.71
## - NumCompaniesWorked           1    679.51  767.51
## - EnvironmentSatisfaction      3    689.03  773.03
## - JobSatisfaction              3    689.85  773.85
## - BusinessTravel               2    687.97  773.97
## - MaritalStatus                2    693.88  779.88
## - OverTime                     1    756.17  844.17
##
## Step:   AIC=749.08
## Attrition ~ Age + BusinessTravel + DailyRate + DistanceFromHome +
##     EnvironmentSatisfaction + Gender + JobInvolvement + JobLevel +
##     JobRole + JobSatisfaction + MaritalStatus + NumCompaniesWorked +
##     OverTime + RelationshipSatisfaction + TotalWorkingYears +
##     WorkLifeBalance + YearsAtCompany + YearsInCurrentRole +
##     YearsSinceLastPromotion +
##     YearsWithCurrManager
##
##                                Df  Deviance     AIC
## - JobLevel                      1    670.64  748.64
## <none>                               669.08  749.08
## - DailyRate                     1    671.70  749.70
## + EducationField                5    659.88  749.88
## + TrainingTimesLastYear         1    668.03  750.03
## - Age                           1    672.25  750.25
## + MonthlyIncome                 1    668.77  750.77
## + PercentSalaryHike             1    668.79  750.79
## + HourlyRate                    1    668.95  750.95
## + PerformanceRating             1    668.97  750.97
## + StockOptionLevel              1    669.05  751.05
## + MonthlyRate                   1    669.08  751.08
## - YearsAtCompany                1    673.39  751.39
## + Department                    2    668.38  752.38
## - YearsWithCurrManager          1    675.43  753.43
## - RelationshipSatisfaction      3    679.64  753.64
## - Gender                        1    675.73  753.73
## + Education                     4    666.00  754.00
## - YearsInCurrentRole            1    676.50  754.50
## - TotalWorkingYears             1    679.26  757.26
## - WorkLifeBalance               3    683.27  757.27
## - DistanceFromHome              1    683.07  761.07
## - JobInvolvement                3    687.50  761.50
## - YearsSinceLastPromotion       1    684.38  762.38
## - JobRole                       8    702.32  766.32
## - NumCompaniesWorked            1    688.77  766.77
## - EnvironmentSatisfaction       3    698.95  772.95
## - JobSatisfaction               3    698.99  772.99
```

```
## - BusinessTravel               2    697.82  773.82
## - MaritalStatus                2    704.21  780.21
## - OverTime                     1    767.17  845.17
##
## Step:   AIC=748.64
## Attrition ~ Age + BusinessTravel + DailyRate + DistanceFromHome +
##      EnvironmentSatisfaction + Gender + JobInvolvement + JobRole +
##      JobSatisfaction + MaritalStatus + NumCompaniesWorked + OverTime +
##      RelationshipSatisfaction + TotalWorkingYears + WorkLifeBalance +
##      YearsAtCompany + YearsInCurrentRole + YearsSinceLastPromotion +
##      YearsWithCurrManager
##
##                                Df Deviance     AIC
## <none>                              670.64  748.64
## + JobLevel                     1    669.08  749.08
## + MonthlyIncome                1    669.09  749.09
## + EducationField               5    661.25  749.25
## - DailyRate                    1    673.35  749.35
## + TrainingTimesLastYear        1    669.51  749.51
## - Age                          1    673.97  749.97
## + PercentSalaryHike            1    670.35  750.35
## + PerformanceRating            1    670.52  750.52
## + HourlyRate                   1    670.54  750.54
## + StockOptionLevel             1    670.60  750.60
## + MonthlyRate                  1    670.63  750.63
## - YearsAtCompany               1    675.70  751.70
## + Department                   2    669.94  751.94
## - YearsWithCurrManager         1    677.23  753.23
## + Education                    4    667.34  753.34
## - Gender                       1    677.52  753.52
## - RelationshipSatisfaction     3    681.60  753.60
## - YearsInCurrentRole           1    678.28  754.28
## - TotalWorkingYears            1    679.26  755.26
## - WorkLifeBalance              3    684.41  756.41
## - DistanceFromHome             1    685.47  761.47
## - JobInvolvement               3    689.60  761.60
## - YearsSinceLastPromotion      1    685.84  761.84
## - JobRole                      8    702.91  764.91
## - NumCompaniesWorked           1    690.27  766.27
## - JobSatisfaction              3    700.18  772.18
## - EnvironmentSatisfaction      3    700.38  772.38
## - BusinessTravel               2    699.56  773.56
## - MaritalStatus                2    705.86  779.86
## - OverTime                     1    768.29  844.29
```

```r
print(stepwise_model)
```

```
##
## Call:  glm(formula = Attrition ~ Age + BusinessTravel + DailyRate +
##      DistanceFromHome + EnvironmentSatisfaction + Gender + JobInvolvement +
##      JobRole + JobSatisfaction + MaritalStatus + NumCompaniesWorked +
##      OverTime + RelationshipSatisfaction + TotalWorkingYears +
##      WorkLifeBalance + YearsAtCompany + YearsInCurrentRole +
##   YearsSinceLastPromotion +
```

```
##          YearsWithCurrManager, family = binomial, data = train_data)
##
## Coefficients:
##                        (Intercept)                                Age
##                         -0.5352665                         -0.0280563
##    BusinessTravelTravel_Frequently       BusinessTravelTravel_Rarely
##                          2.3384668                          1.4202249
##                          DailyRate                   DistanceFromHome
##                         -0.0004094                          0.0473885
##       EnvironmentSatisfactionMedium       EnvironmentSatisfactionHigh
##                         -1.3046390                         -1.2695789
##   EnvironmentSatisfactionVery High                         GenderMale
##                         -1.4124949                          0.5445888
##              JobInvolvementMedium                  JobInvolvementHigh
##                         -0.8301588                         -1.1763826
##           JobInvolvementVery High              JobRoleHuman Resources
##                         -2.2147452                          1.3533823
##        JobRoleLaboratory Technician                      JobRoleManager
##                          1.4430516                          0.1743846
##       JobRoleManufacturing Director          JobRoleResearch Director
##                          0.1344172                         -0.5957621
##         JobRoleResearch Scientist          JobRoleSales Executive
##                          0.6265961                          1.2209278
##        JobRoleSales Representative          JobSatisfactionMedium
##                          2.1118910                         -0.7280826
##              JobSatisfactionHigh            JobSatisfactionVery High
##                         -0.5725076                         -1.5399987
##             MaritalStatusMarried                  MaritalStatusSingle
##                          0.5716821                          1.5826218
##               NumCompaniesWorked                          OverTimeYes
##                          0.1966050                          2.0726675
##     RelationshipSatisfactionMedium    RelationshipSatisfactionHigh
##                         -0.7591370                         -0.7972933
## RelationshipSatisfactionVery High                   TotalWorkingYears
##                         -0.8733882                         -0.0915575
##               WorkLifeBalanceGood               WorkLifeBalanceBetter
##                         -0.7522929                         -1.2999740
##               WorkLifeBalanceBest                       YearsAtCompany
##                         -0.9359163                          0.1048887
##               YearsInCurrentRole             YearsSinceLastPromotion
##                         -0.1430508                          0.1862112
##             YearsWithCurrManager
##                         -0.1364658
##
## Degrees of Freedom: 1176 Total (i.e. Null);   1138 Residual
## Null Deviance:        1041
## Residual Deviance: 670.6      AIC: 748.6
```
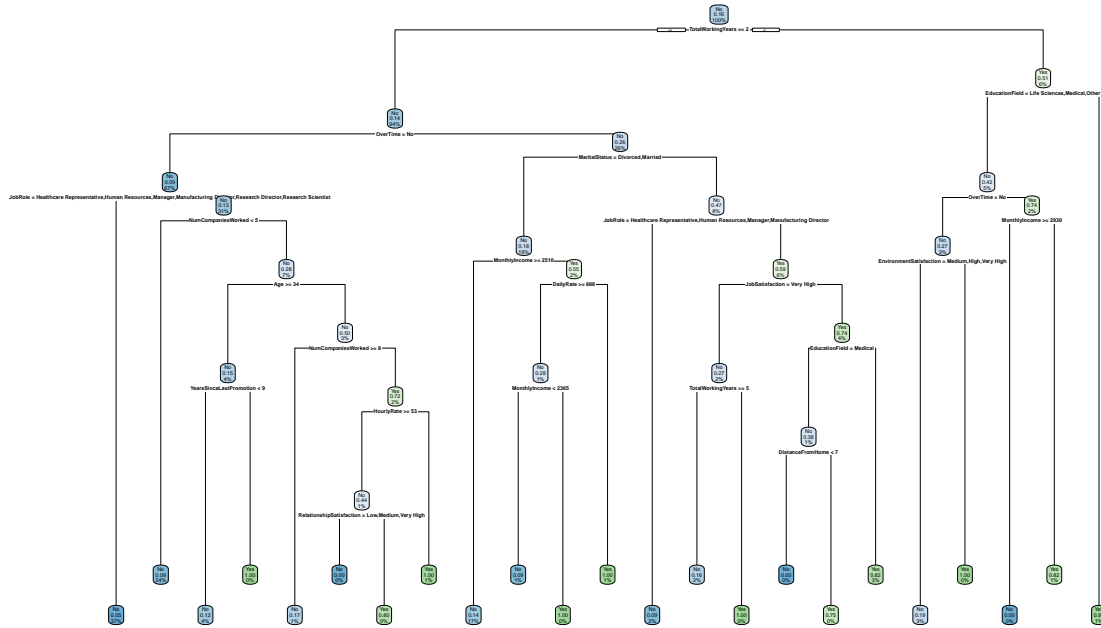
```r
basic_tree <- rpart(Attrition ~ ., data = train_data, method = "class",
    control = rpart.control(minsplit = 2, minbucket = 1, maxsurrogate = 0))
rpart.plot(basic_tree, main = "Basic Classification Tree")
```

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

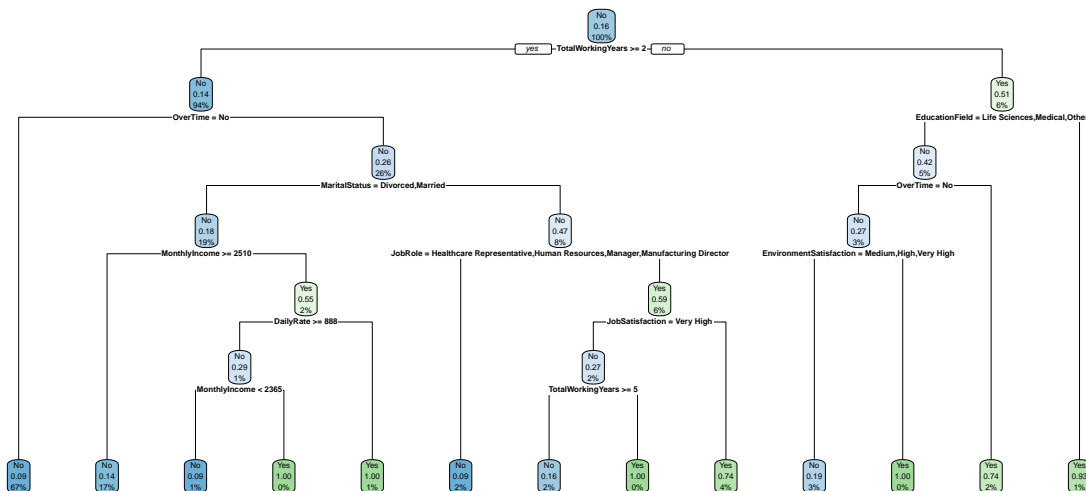**Basic Classification Tree**


```
full_tree <- rpart(Attrition ~ ., data = train_data, method = "class", cp = 0,
    control = rpart.control(minsplit = 2, minbucket = 1, maxsurrogate = 0))
best_cp <- full_tree$cptable[which.min(full_tree$cptable[,"xerror"]),"CP"]
pruned_tree <- prune(full_tree, cp = best_cp)
rpart.plot(pruned_tree, main = "Pruned Classification Tree")
```

Pruned Classification Tree


```
print(pruned_tree)

## n= 1177
##
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
##  1) root 1177 190 No (0.83857264 0.16142736)
##    2) TotalWorkingYears>=1.5 1103 152 No (0.86219402 0.13780598)
##      4) OverTime=No 793   70 No (0.91172762 0.08827238) *
##      5) OverTime=Yes 310   82 No (0.73548387 0.26451613)
##       10) MaritalStatus=Divorced,Married 218   39 No (0.82110092 0.17889908)
##         20) MonthlyIncome>=2510 196   27 No (0.86224490 0.13775510) *
##         21) MonthlyIncome< 2510 22   10 Yes (0.45454545 0.54545455)
##           42) DailyRate>=888 14    4 No (0.71428571 0.28571429)
##             84) MonthlyIncome< 2364.5 11    1 No (0.90909091 0.09090909) *
##             85) MonthlyIncome>=2364.5 3    0 Yes (0.00000000 1.00000000) *
##           43) DailyRate< 888 8    0 Yes (0.00000000 1.00000000) *
##       11) MaritalStatus=Single 92   43 No (0.53260870 0.46739130)
##         22) JobRole=Healthcare Representative,Human Resources,Manager,
##   Manufacturing Director 23    2 No (0.91304348 0.08695652) *
##         23) JobRole=Laboratory Technician,Research Director,Research
##   Scientist,Sales Executive,Sales Representative 69   28 Yes (0.40579710
##   0.59420290)
##           46) JobSatisfaction=Very High 22    6 No (0.72727273 0.27272727)
##             92) TotalWorkingYears>=4.5 19    3 No (0.84210526 0.15789474) *
##             93) TotalWorkingYears< 4.5 3    0 Yes (0.00000000 1.00000000) *
```

```
##           47) JobSatisfaction=Low,Medium,High 47   12 Yes (0.25531915
   0.74468085) *
##    3) TotalWorkingYears< 1.5 74   36 Yes (0.48648649 0.51351351)
##      6) EducationField=Life Sciences,Medical,Other 60   25 No (0.58333333
   0.41666667)
##        12) OverTime=No 41   11 No (0.73170732 0.26829268)
##          24) EnvironmentSatisfaction=Medium,High,Very High 37    7 No
   (0.81081081 0.18918919) *
##          25) EnvironmentSatisfaction=Low 4    0 Yes (0.00000000 1.00000000) *
##        13) OverTime=Yes 19    5 Yes (0.26315789 0.73684211) *
##      7) EducationField=Human Resources,Marketing,Technical Degree 14    1
   Yes (0.07142857 0.92857143) *
```

Model Evaluation

```r
create_confusion_matrices <- function(model, test_data, thresholds, model_type
   ) {
  results <- list()
  if (model_type == "logistic") {
    predictions <- predict(model, newdata = test_data, type = "response")
  } else {
    predictions <- predict(model, newdata = test_data, type = "prob")[,2]
  }
  for (threshold in thresholds) {
    predicted_classes <- factor(ifelse(predictions > threshold, "Yes", "No"),
        levels = c("No", "Yes"))
    confusion <- confusionMatrix(predicted_classes, test_data$Attrition,
        positive = "Yes")
    results[[paste0("Threshold_", threshold)]] <- confusion
  }
  return(results)
}

evaluate_model_performance <- function(confusion) {
  list(
    Accuracy = confusion$overall["Accuracy"],
    BalancedAccuracy = confusion$byClass["Balanced Accuracy"]
  )
}

thresholds <- seq(0.05, 0.95, by = 0.05)

confusion_matrices_lr <- create_confusion_matrices(stepwise_model, test_data,
   thresholds, "logistic")
performance_lr <- do.call(rbind, lapply(names(confusion_matrices_lr), function
   (x) {
  metrics <- evaluate_model_performance(confusion_matrices_lr[[x]])
  data.frame(
    Model = "Logistic Regression",
    Threshold = gsub("Threshold_", "", x),
    Accuracy = metrics$Accuracy,
    BalancedAccuracy = metrics$BalancedAccuracy
```

```
  )
}))

confusion_matrices_pruned <- create_confusion_matrices(pruned_tree, test_data,
    thresholds, "tree")
performance_pruned <- do.call(rbind, lapply(names(confusion_matrices_pruned),
   function(x) {
  metrics <- evaluate_model_performance(confusion_matrices_pruned[[x]])
  data.frame(
    Model = "Pruned Tree",
    Threshold = gsub("Threshold_", "", x),
    Accuracy = metrics$Accuracy,
    BalancedAccuracy = metrics$BalancedAccuracy
  )
}))

performance_combined <- rbind(performance_lr, performance_pruned)
write.xlsx(performance_combined, "/Users/hunterberberich/Desktop/School Stuff!
    /Data/ModelPerformanceCombined.xlsx")

print(performance_combined)
```

```
##                          Model Threshold   Accuracy  BalancedAccuracy
## Accuracy      Logistic Regression    0.05  0.5972696         0.7085279
## Accuracy1     Logistic Regression     0.1  0.7201365         0.7730929
## Accuracy2     Logistic Regression    0.15  0.7542662         0.7503892
## Accuracy3     Logistic Regression     0.2  0.7986348         0.7596004
## Accuracy4     Logistic Regression    0.25  0.8191126         0.7373724
## Accuracy5     Logistic Regression     0.3  0.8361775         0.7217177
## Accuracy6     Logistic Regression    0.35  0.8464164         0.7106037
## Accuracy7     Logistic Regression     0.4  0.8600683         0.7015222
## Accuracy8     Logistic Regression    0.45  0.8703072         0.6990140
## Accuracy9     Logistic Regression     0.5  0.8703072         0.6818025
## Accuracy10    Logistic Regression    0.55  0.8839590         0.6813268
## Accuracy11    Logistic Regression     0.6  0.8839590         0.6727210
## Accuracy12    Logistic Regression    0.65  0.8805461         0.6620827
## Accuracy13    Logistic Regression     0.7  0.8771331         0.6342328
## Accuracy14    Logistic Regression    0.75  0.8737201         0.6063830
## Accuracy15    Logistic Regression     0.8  0.8634812         0.5744681
## Accuracy16    Logistic Regression    0.85  0.8600683         0.5638298
## Accuracy17    Logistic Regression     0.9  0.8532423         0.5425532
## Accuracy18    Logistic Regression    0.95  0.8464164         0.5212766
## Accuracy19             Pruned Tree    0.05  0.1604096         0.5000000
## Accuracy110            Pruned Tree     0.1  0.7337884         0.6865594
## Accuracy21             Pruned Tree    0.15  0.8259386         0.6639855
## Accuracy31             Pruned Tree     0.2  0.8225256         0.6017125
## Accuracy41             Pruned Tree    0.25  0.8225256         0.6017125
## Accuracy51             Pruned Tree     0.3  0.8225256         0.6017125
## Accuracy61             Pruned Tree    0.35  0.8225256         0.6017125
## Accuracy71             Pruned Tree     0.4  0.8225256         0.6017125
## Accuracy81             Pruned Tree    0.45  0.8225256         0.6017125
## Accuracy91             Pruned Tree     0.5  0.8225256         0.6017125
## Accuracy101            Pruned Tree    0.55  0.8225256         0.6017125
## Accuracy111            Pruned Tree     0.6  0.8225256         0.6017125
```

```
## Accuracy121            Pruned Tree          0.65  0.8225256           0.6017125
## Accuracy131            Pruned Tree           0.7  0.8225256           0.6017125
## Accuracy141            Pruned Tree          0.75  0.8327645           0.5389638
## Accuracy151            Pruned Tree           0.8  0.8327645           0.5389638
## Accuracy161            Pruned Tree          0.85  0.8327645           0.5389638
## Accuracy171            Pruned Tree           0.9  0.8327645           0.5389638
## Accuracy181            Pruned Tree          0.95  0.8395904           0.5430289
```
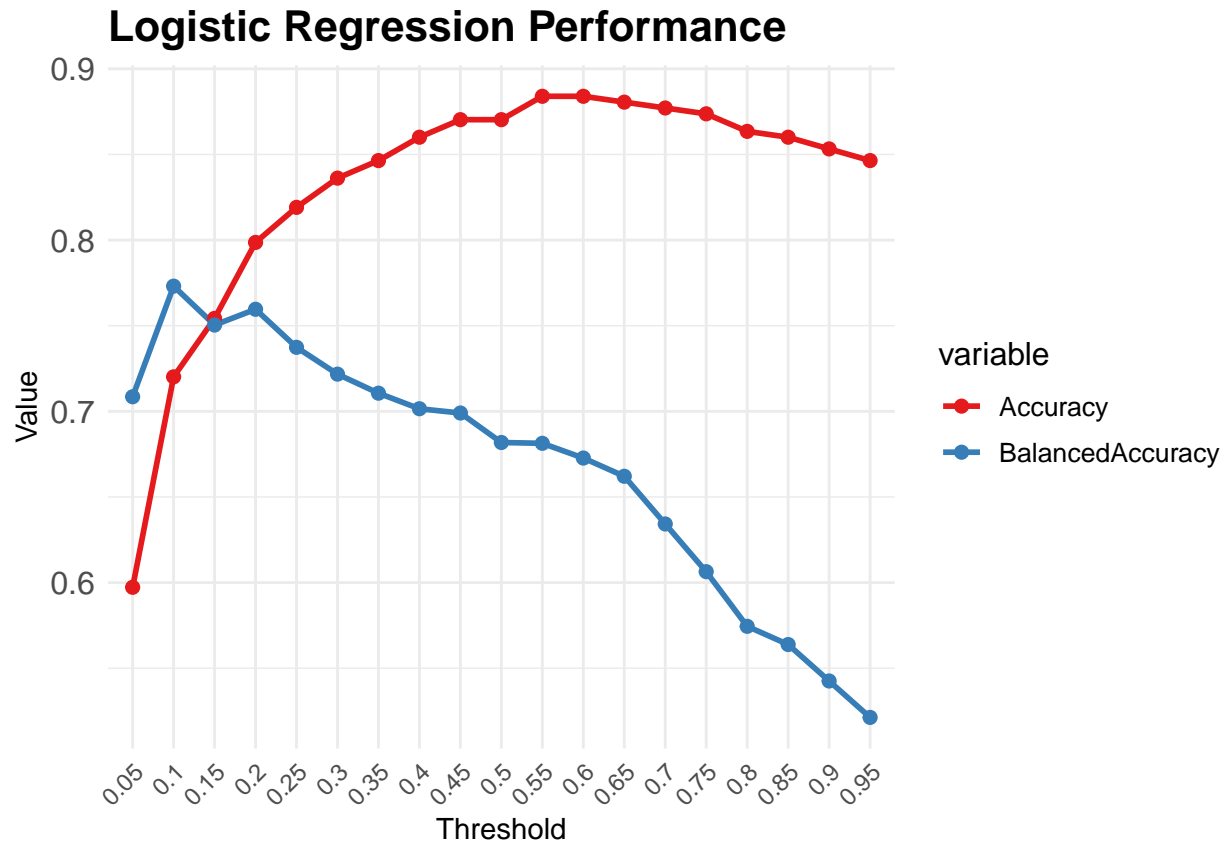
```r
options(repr.plot.width = 12, repr.plot.height = 6)

performance_lr_long <- melt(performance_lr, id.vars = c("Model", "Threshold"),
    measure.vars = c("Accuracy", "BalancedAccuracy"))
ggplot(performance_lr_long, aes(x = as.factor(Threshold), y = value, color =
    variable, group = variable)) +
  geom_line(size = 1.0) +
  geom_point(size = 2) +
  theme_minimal() +
  labs(title = "Logistic Regression Performance", x = "Threshold", y = "Value"
    ) +
  scale_color_brewer(palette = "Set1") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 9),
        axis.text.y = element_text(size = 12),
        plot.title = element_text(size = 16, face = "bold"),
        legend.title = element_text(size = 12),
        legend.text = element_text(size = 10))
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

Logistic Regression Performance


```
ggsave("Logistic_Regression_Performance_Metrics.png", width = 12, height = 6)

performance_tree_long <- melt(performance_pruned, id.vars = c("Model", "
    Threshold"), measure.vars = c("Accuracy", "BalancedAccuracy"))
ggplot(performance_tree_long, aes(x = as.factor(Threshold), y = value, color =
     variable, group = variable)) +
  geom_line(size = 1.0) +
  geom_point(size = 2) +
  theme_minimal() +
  labs(title = "Pruned Tree Performance Metrics", x = "Threshold", y = "Value"
    ) +
  scale_color_brewer(palette = "Set1") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 12),
        axis.text.y = element_text(size = 12),
        plot.title = element_text(size = 16, face = "bold"),
        legend.title = element_text(size = 12),
        legend.text = element_text(size = 10))
```

Pruned Tree Performance Metrics
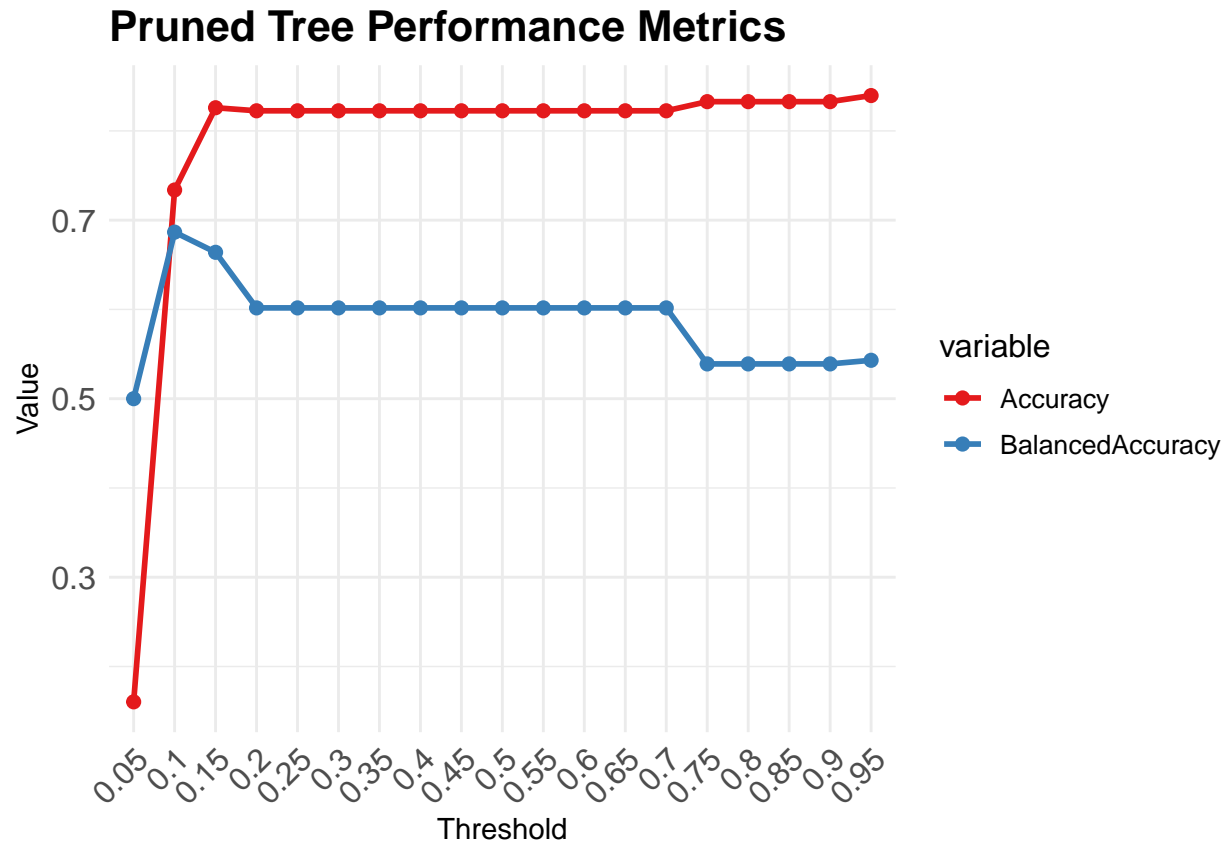

```r
plot_roc <- function(model, test_data, model_type) {
  if (model_type == "logistic") {
    predictions <- predict(model, newdata = test_data, type = "response")
  } else {
    predictions <- predict(model, newdata = test_data, type = "prob")[,2]
  }
  roc_obj <- roc(test_data$Attrition, predictions, levels = rev(levels(test_
      data$Attrition)))
  plot(roc_obj, main = paste(model_type, "ROC␣Curve"), col = "blue")
  return(roc_obj)
}

roc_lr <- plot_roc(stepwise_model, test_data, "logistic")

## Setting direction: controls > cases
```
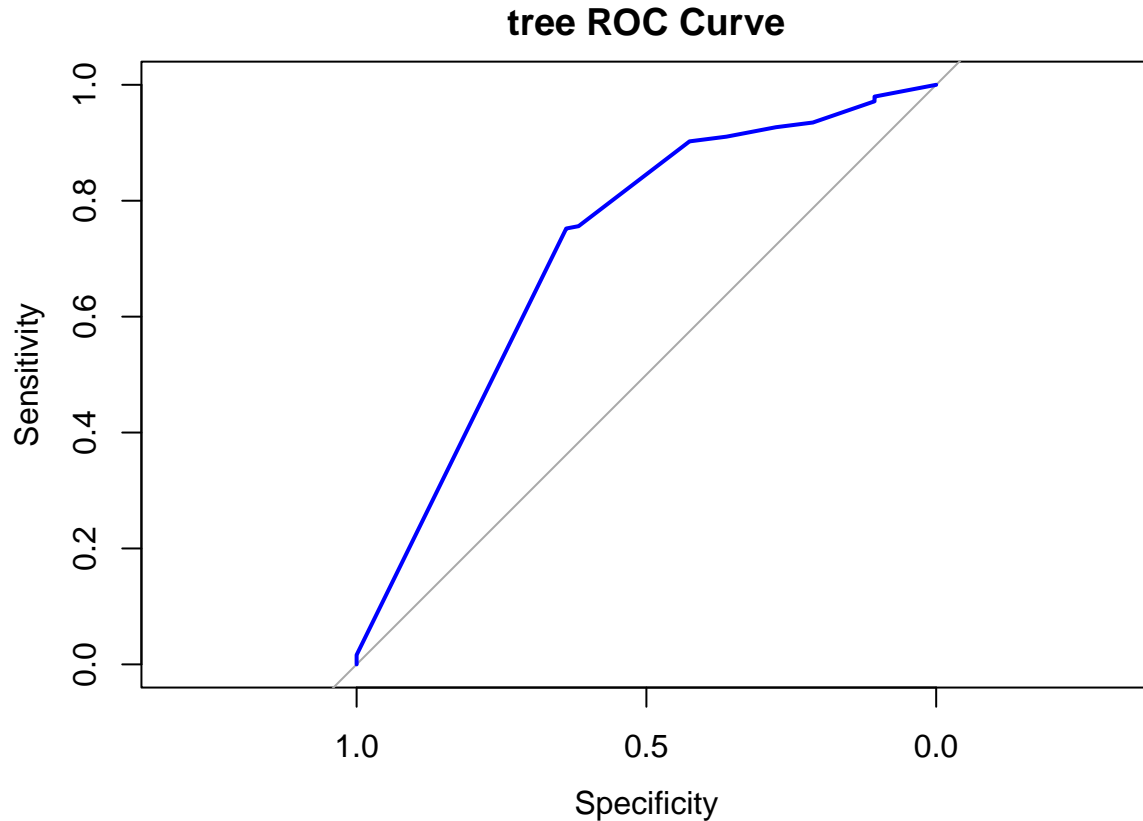
logistic ROC Curve
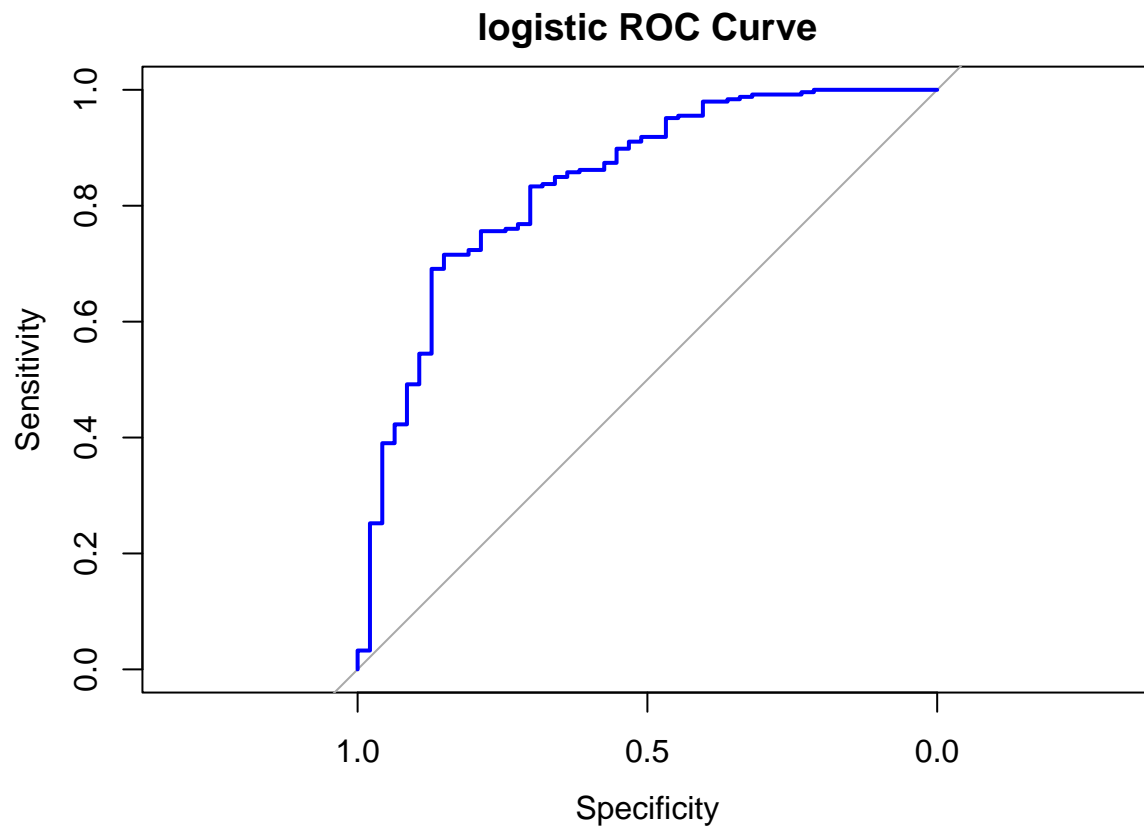

```
roc_pruned <- plot_roc(pruned_tree, test_data, "tree")

## Setting direction: controls > cases
```

tree ROC Curve

Cost Analysis

```
calculate_cost_savings <- function(confusion_matrix, total_leaving_employees,
    replacement_cost, raise_cost) {
  false_negatives <- confusion_matrix$table[1, 2]
  false_positives <- confusion_matrix$table[2, 1]

  cost_without_model <- total_leaving_employees * replacement_cost

  cost_with_model <- (false_negatives * replacement_cost) + (false_positives *
      raise_cost)

  cost_savings <- cost_without_model - cost_with_model

  return(list(cost_without_model = cost_without_model, cost_with_model = cost_
      with_model, cost_savings = cost_savings))
}

annual_salary <- 78035
replacement_cost <- 1.5 * annual_salary
raise_cost <- 0.1 * annual_salary

cost_savings_results_lr <- lapply(names(confusion_matrices_lr), function(
    threshold) {
```

```r
  confusion_matrix <- confusion_matrices_lr [[threshold]]
  total_leaving_employees <- sum(test_data$Attrition == "Yes")
  cost_savings <- calculate_cost_savings(confusion_matrix, total_leaving_
      employees, replacement_cost, raise_cost)
  return(data.frame(Threshold = as.numeric(gsub("Threshold_", "", threshold)),
                    Model = "Logistic Regression",
                    Cost_Savings = cost_savings$cost_savings))
})

cost_savings_results_pruned <- lapply(names(confusion_matrices_pruned),
    function(threshold) {
  confusion_matrix <- confusion_matrices_pruned [[threshold]]
  total_leaving_employees <- sum(test_data$Attrition == "Yes")
  cost_savings <- calculate_cost_savings(confusion_matrix, total_leaving_
      employees, replacement_cost, raise_cost)
  return(data.frame(Threshold = as.numeric(gsub("Threshold_", "", threshold)),
                    Model = "Pruned Tree",
                    Cost_Savings = cost_savings$cost_savings))
})

cost_savings_df_lr <- do.call(rbind, cost_savings_results_lr)
cost_savings_df_pruned <- do.call(rbind, cost_savings_results_pruned)
cost_savings_df <- rbind(cost_savings_df_lr, cost_savings_df_pruned)

ggplot(cost_savings_df, aes(x = Threshold, y = Cost_Savings, color = Model,
    group = Model)) +
  geom_line(size = 1.0) +
  geom_point(size = 2) +
  scale_x_continuous(breaks = seq(0, 1, by = 0.05)) +
  theme_minimal() +
  labs(title = "Cost Savings at Different Thresholds", x = "Threshold", y = "
      Cost Savings") +
  scale_color_brewer(palette = "Set1") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 9),
        axis.text.y = element_text(size = 12),
        plot.title = element_text(size = 16, face = "bold"),
        legend.title = element_text(size = 12),
        legend.text = element_text(size = 10))
```

Cost Savings at Different Thresholds