Employee Attrition Prediction

Inushi Kashmira

2025-09-01

#install.packages(c("tidyverse","GGally","caret","pROC"))  
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.4.3

## Warning: package 'ggplot2' was built under R version 4.4.3

## Warning: package 'tidyr' was built under R version 4.4.3

## Warning: package 'readr' was built under R version 4.4.3

## Warning: package 'purrr' was built under R version 4.4.3

## Warning: package 'dplyr' was built under R version 4.4.3

## Warning: package 'forcats' was built under R version 4.4.3

## Warning: package 'lubridate' was built under R version 4.4.3

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.4   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(GGally)

## Warning: package 'GGally' was built under R version 4.4.3

library(caret)

## Warning: package 'caret' was built under R version 4.4.3

## Loading required package: lattice  
##   
## Attaching package: 'caret'  
##   
## The following object is masked from 'package:purrr':  
##   
## lift

library(pROC)

## Warning: package 'pROC' was built under R version 4.4.3

## Type 'citation("pROC")' for a citation.  
##   
## Attaching package: 'pROC'  
##   
## The following objects are masked from 'package:stats':  
##   
## cov, smooth, var

data<-read.csv("D://project1//WA\_Fn-UseC\_-HR-Employee-Attrition.csv")  
head(data)

## Age Attrition BusinessTravel DailyRate Department  
## 1 41 Yes Travel\_Rarely 1102 Sales  
## 2 49 No Travel\_Frequently 279 Research & Development  
## 3 37 Yes Travel\_Rarely 1373 Research & Development  
## 4 33 No Travel\_Frequently 1392 Research & Development  
## 5 27 No Travel\_Rarely 591 Research & Development  
## 6 32 No Travel\_Frequently 1005 Research & Development  
## DistanceFromHome Education EducationField EmployeeCount EmployeeNumber  
## 1 1 2 Life Sciences 1 1  
## 2 8 1 Life Sciences 1 2  
## 3 2 2 Other 1 4  
## 4 3 4 Life Sciences 1 5  
## 5 2 1 Medical 1 7  
## 6 2 2 Life Sciences 1 8  
## EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel  
## 1 2 Female 94 3 2  
## 2 3 Male 61 2 2  
## 3 4 Male 92 2 1  
## 4 4 Female 56 3 1  
## 5 1 Male 40 3 1  
## 6 4 Male 79 3 1  
## JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate  
## 1 Sales Executive 4 Single 5993 19479  
## 2 Research Scientist 2 Married 5130 24907  
## 3 Laboratory Technician 3 Single 2090 2396  
## 4 Research Scientist 3 Married 2909 23159  
## 5 Laboratory Technician 2 Married 3468 16632  
## 6 Laboratory Technician 4 Single 3068 11864  
## NumCompaniesWorked Over18 OverTime PercentSalaryHike PerformanceRating  
## 1 8 Y Yes 11 3  
## 2 1 Y No 23 4  
## 3 6 Y Yes 15 3  
## 4 1 Y Yes 11 3  
## 5 9 Y No 12 3  
## 6 0 Y No 13 3  
## RelationshipSatisfaction StandardHours StockOptionLevel TotalWorkingYears  
## 1 1 80 0 8  
## 2 4 80 1 10  
## 3 2 80 0 7  
## 4 3 80 0 8  
## 5 4 80 1 6  
## 6 3 80 0 8  
## TrainingTimesLastYear WorkLifeBalance YearsAtCompany YearsInCurrentRole  
## 1 0 1 6 4  
## 2 3 3 10 7  
## 3 3 3 0 0  
## 4 3 3 8 7  
## 5 3 3 2 2  
## 6 2 2 7 7  
## YearsSinceLastPromotion YearsWithCurrManager  
## 1 0 5  
## 2 1 7  
## 3 0 0  
## 4 3 0  
## 5 2 2  
## 6 3 6

str(data)

## 'data.frame': 1470 obs. of 35 variables:  
## $ Age : int 41 49 37 33 27 32 59 30 38 36 ...  
## $ Attrition : chr "Yes" "No" "Yes" "No" ...  
## $ BusinessTravel : chr "Travel\_Rarely" "Travel\_Frequently" "Travel\_Rarely" "Travel\_Frequently" ...  
## $ DailyRate : int 1102 279 1373 1392 591 1005 1324 1358 216 1299 ...  
## $ Department : chr "Sales" "Research & Development" "Research & Development" "Research & Development" ...  
## $ 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 : chr "Life Sciences" "Life Sciences" "Other" "Life Sciences" ...  
## $ EmployeeCount : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ EmployeeNumber : int 1 2 4 5 7 8 10 11 12 13 ...  
## $ EnvironmentSatisfaction : int 2 3 4 4 1 4 3 4 4 3 ...  
## $ Gender : chr "Female" "Male" "Male" "Female" ...  
## $ 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 : chr "Sales Executive" "Research Scientist" "Laboratory Technician" "Research Scientist" ...  
## $ JobSatisfaction : int 4 2 3 3 2 4 1 3 3 3 ...  
## $ MaritalStatus : chr "Single" "Married" "Single" "Married" ...  
## $ 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 ...  
## $ Over18 : chr "Y" "Y" "Y" "Y" ...  
## $ OverTime : chr "Yes" "No" "Yes" "Yes" ...  
## $ 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 ...  
## $ StandardHours : int 80 80 80 80 80 80 80 80 80 80 ...  
## $ 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 ...

summary(data)

## Age Attrition BusinessTravel DailyRate   
## Min. :18.00 Length:1470 Length:1470 Min. : 102.0   
## 1st Qu.:30.00 Class :character Class :character 1st Qu.: 465.0   
## Median :36.00 Mode :character Mode :character 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 EducationField   
## Length:1470 Min. : 1.000 Min. :1.000 Length:1470   
## Class :character 1st Qu.: 2.000 1st Qu.:2.000 Class :character   
## Mode :character Median : 7.000 Median :3.000 Mode :character   
## Mean : 9.193 Mean :2.913   
## 3rd Qu.:14.000 3rd Qu.:4.000   
## Max. :29.000 Max. :5.000   
## EmployeeCount EmployeeNumber EnvironmentSatisfaction Gender   
## Min. :1 Min. : 1.0 Min. :1.000 Length:1470   
## 1st Qu.:1 1st Qu.: 491.2 1st Qu.:2.000 Class :character   
## Median :1 Median :1020.5 Median :3.000 Mode :character   
## Mean :1 Mean :1024.9 Mean :2.722   
## 3rd Qu.:1 3rd Qu.:1555.8 3rd Qu.:4.000   
## Max. :1 Max. :2068.0 Max. :4.000   
## HourlyRate JobInvolvement JobLevel JobRole   
## Min. : 30.00 Min. :1.00 Min. :1.000 Length:1470   
## 1st Qu.: 48.00 1st Qu.:2.00 1st Qu.:1.000 Class :character   
## Median : 66.00 Median :3.00 Median :2.000 Mode :character   
## 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   
## JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate   
## Min. :1.000 Length:1470 Min. : 1009 Min. : 2094   
## 1st Qu.:2.000 Class :character 1st Qu.: 2911 1st Qu.: 8047   
## Median :3.000 Mode :character Median : 4919 Median :14236   
## Mean :2.729 Mean : 6503 Mean :14313   
## 3rd Qu.:4.000 3rd Qu.: 8379 3rd Qu.:20462   
## Max. :4.000 Max. :19999 Max. :26999   
## NumCompaniesWorked Over18 OverTime PercentSalaryHike  
## Min. :0.000 Length:1470 Length:1470 Min. :11.00   
## 1st Qu.:1.000 Class :character Class :character 1st Qu.:12.00   
## Median :2.000 Mode :character Mode :character Median :14.00   
## Mean :2.693 Mean :15.21   
## 3rd Qu.:4.000 3rd Qu.:18.00   
## Max. :9.000 Max. :25.00   
## PerformanceRating RelationshipSatisfaction StandardHours StockOptionLevel  
## Min. :3.000 Min. :1.000 Min. :80 Min. :0.0000   
## 1st Qu.:3.000 1st Qu.:2.000 1st Qu.:80 1st Qu.:0.0000   
## Median :3.000 Median :3.000 Median :80 Median :1.0000   
## Mean :3.154 Mean :2.712 Mean :80 Mean :0.7939   
## 3rd Qu.:3.000 3rd Qu.:4.000 3rd Qu.:80 3rd Qu.:1.0000   
## Max. :4.000 Max. :4.000 Max. :80 Max. :3.0000   
## TotalWorkingYears TrainingTimesLastYear WorkLifeBalance YearsAtCompany   
## Min. : 0.00 Min. :0.000 Min. :1.000 Min. : 0.000   
## 1st Qu.: 6.00 1st Qu.:2.000 1st Qu.:2.000 1st Qu.: 3.000   
## Median :10.00 Median :3.000 Median :3.000 Median : 5.000   
## Mean :11.28 Mean :2.799 Mean :2.761 Mean : 7.008   
## 3rd Qu.:15.00 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.: 9.000   
## Max. :40.00 Max. :6.000 Max. :4.000 Max. :40.000   
## YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager  
## Min. : 0.000 Min. : 0.000 Min. : 0.000   
## 1st Qu.: 2.000 1st Qu.: 0.000 1st Qu.: 2.000   
## Median : 3.000 Median : 1.000 Median : 3.000   
## Mean : 4.229 Mean : 2.188 Mean : 4.123   
## 3rd Qu.: 7.000 3rd Qu.: 3.000 3rd Qu.: 7.000   
## Max. :18.000 Max. :15.000 Max. :17.000

# Convert categorical variables to factors  
categorical\_vars <- c("BusinessTravel","Department","EducationField","Gender",  
 "JobRole","MaritalStatus","OverTime","Attrition")  
  
data[categorical\_vars] <- lapply(data[categorical\_vars], as.factor)  
  
# Encode target variable: Attrition Yes = 1, No = 0  
data$Attrition <- ifelse(data$Attrition=="Yes", 1, 0)  
  
# Check missing values  
sum(is.na(data))

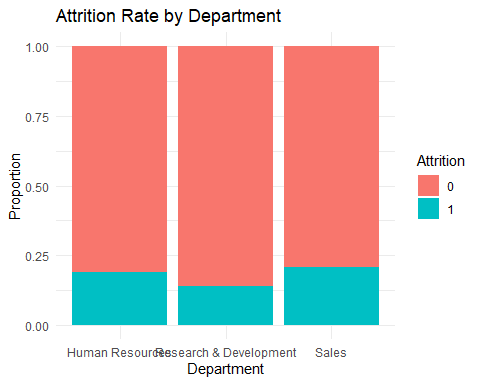
## [1] 0

install.packages("ggplot2")

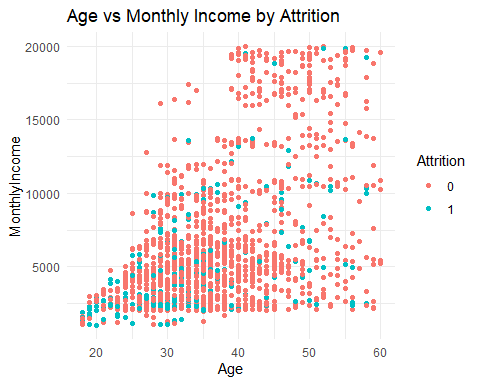
## Warning: package 'ggplot2' is in use and will not be installed

library(ggplot2)

# Attrition rate by Department  
ggplot(data, aes(x=Department, fill=factor(Attrition))) +  
 geom\_bar(position="fill") +  
 labs(title="Attrition Rate by Department", y="Proportion", fill="Attrition") +  
 theme\_minimal()



# Scatter plot of Age vs MonthlyIncome colored by Attrition  
ggplot(data, aes(x=Age, y=MonthlyIncome, color=factor(Attrition))) +  
 geom\_point() +  
 labs(title="Age vs Monthly Income by Attrition", color="Attrition") +  
 theme\_minimal()



set.seed(123)  
trainIndex <- createDataPartition(data$Attrition, p=0.7, list=FALSE)  
train\_data <- data[trainIndex,]  
test\_data <- data[-trainIndex,]

# Build logistic regression model  
model <- glm(Attrition ~ Age + MonthlyIncome + JobSatisfaction +   
 YearsAtCompany + OverTime + DistanceFromHome + JobLevel,  
 data=train\_data, family=binomial)  
  
summary(model)

##   
## Call:  
## glm(formula = Attrition ~ Age + MonthlyIncome + JobSatisfaction +   
## YearsAtCompany + OverTime + DistanceFromHome + JobLevel,   
## family = binomial, data = train\_data)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 8.364e-01 4.825e-01 1.733 0.083029 .   
## Age -4.910e-02 1.297e-02 -3.786 0.000153 \*\*\*  
## MonthlyIncome -8.117e-05 6.956e-05 -1.167 0.243288   
## JobSatisfaction -2.953e-01 8.186e-02 -3.607 0.000309 \*\*\*  
## YearsAtCompany -2.332e-02 2.237e-02 -1.042 0.297225   
## OverTimeYes 1.341e+00 1.876e-01 7.149 8.73e-13 \*\*\*  
## DistanceFromHome 2.471e-02 1.121e-02 2.204 0.027508 \*   
## JobLevel -7.428e-02 2.797e-01 -0.266 0.790532   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 896.03 on 1028 degrees of freedom  
## Residual deviance: 764.88 on 1021 degrees of freedom  
## AIC: 780.88  
##   
## Number of Fisher Scoring iterations: 5

# Predict probabilities on test set  
probabilities <- predict(model, newdata=test\_data, type="response")  
probabilities

## 3 7 14 15 21 22   
## 0.327425093 0.221797137 0.132185613 0.544231642 0.178932190 0.186687632   
## 23 27 30 33 43 47   
## 0.061402806 0.520117917 0.025978861 0.143472119 0.270471884 0.132396006   
## 50 53 54 57 59 60   
## 0.104563485 0.281245071 0.335559881 0.129926851 0.076850617 0.066618719   
## 62 63 65 66 70 73   
## 0.111924344 0.028951442 0.044862533 0.056096991 0.348310867 0.174753520   
## 82 85 87 92 97 101   
## 0.100683816 0.117615071 0.309674804 0.140325371 0.144534235 0.475265705   
## 102 104 107 109 110 113   
## 0.207009082 0.077077205 0.053323411 0.147438394 0.203951967 0.055558366   
## 114 116 118 123 125 131   
## 0.231602917 0.046096527 0.044624266 0.196106455 0.290031870 0.119410647   
## 132 135 136 140 146 147   
## 0.028739488 0.230032813 0.119913906 0.235994332 0.229209795 0.101550085   
## 149 150 154 157 173 175   
## 0.164886526 0.303427254 0.077484444 0.045199961 0.132606794 0.289003100   
## 176 181 182 183 184 187   
## 0.242962189 0.328276955 0.251911379 0.412233039 0.050994822 0.013421488   
## 188 192 198 203 213 214   
## 0.010357890 0.217529526 0.092332737 0.347173754 0.118882570 0.028454087   
## 216 219 226 228 231 233   
## 0.098137189 0.022061751 0.031006717 0.053855556 0.038038913 0.042972779   
## 245 246 249 253 254 257   
## 0.006565780 0.053301873 0.152717883 0.141571478 0.375191068 0.348538589   
## 260 267 268 272 273 274   
## 0.262472484 0.108107349 0.235753124 0.194295761 0.135145868 0.056986729   
## 276 283 284 288 295 296   
## 0.096262332 0.077934918 0.115925008 0.090796222 0.277957015 0.045851820   
## 299 300 301 306 308 313   
## 0.112573073 0.052359358 0.021289815 0.148310225 0.080363400 0.306556830   
## 314 317 322 324 325 333   
## 0.064694763 0.063946601 0.067964207 0.319800008 0.121685475 0.054994274   
## 338 340 342 346 350 351   
## 0.117746558 0.151637505 0.128679038 0.225299237 0.114735178 0.252654199   
## 353 354 356 359 360 361   
## 0.050801328 0.358197482 0.112401607 0.057341369 0.033843265 0.026844954   
## 362 363 364 368 369 371   
## 0.300964424 0.180986022 0.369489894 0.023639512 0.281828834 0.308540784   
## 377 388 390 394 395 399   
## 0.038583811 0.083329782 0.378593165 0.085538765 0.166676763 0.327290167   
## 404 405 406 408 414 417   
## 0.070135463 0.237671377 0.257216752 0.066378525 0.085417883 0.258060891   
## 418 427 428 429 435 438   
## 0.013839619 0.090515279 0.039577353 0.036267578 0.079858125 0.191434407   
## 439 440 442 443 453 454   
## 0.257181081 0.107062575 0.092868719 0.041826726 0.066636230 0.493706428   
## 465 476 483 491 492 495   
## 0.042792004 0.247250963 0.227579927 0.169806901 0.198826482 0.377003144   
## 496 497 498 501 505 506   
## 0.240547892 0.276650827 0.043435508 0.239693873 0.175672930 0.379347478   
## 509 510 511 513 514 518   
## 0.047446246 0.052368878 0.037704090 0.126750664 0.593143053 0.198388776   
## 532 543 546 552 553 556   
## 0.041735412 0.192483466 0.128262554 0.079527166 0.013338281 0.224697915   
## 558 567 568 569 570 572   
## 0.152725284 0.243487085 0.059379386 0.053703644 0.109018277 0.111858188   
## 573 574 580 582 585 587   
## 0.183416939 0.229333396 0.161271955 0.126759182 0.009599687 0.263744922   
## 596 604 607 609 613 614   
## 0.020468985 0.076427698 0.117421550 0.022460384 0.386579704 0.084112785   
## 621 624 629 636 638 641   
## 0.282450711 0.100607633 0.103679125 0.155608394 0.097664862 0.133314776   
## 644 645 652 659 661 662   
## 0.156862475 0.340816882 0.125396800 0.371765783 0.109763468 0.097384253   
## 669 676 684 694 695 701   
## 0.180068567 0.138799704 0.630684855 0.106084824 0.058985512 0.041214381   
## 704 707 716 717 718 720   
## 0.206245504 0.235817816 0.322382120 0.014663101 0.262354428 0.129076226   
## 723 731 732 733 734 735   
## 0.118457121 0.325334025 0.703313321 0.190834536 0.086961779 0.330033459   
## 736 738 739 745 746 748   
## 0.057456405 0.076487431 0.017021415 0.134835043 0.179089339 0.178817320   
## 760 761 762 763 766 768   
## 0.146568565 0.029579660 0.118068100 0.585356420 0.100544288 0.111838816   
## 773 777 782 783 784 795   
## 0.047129792 0.481760183 0.227446754 0.177333500 0.077897990 0.054260363   
## 796 800 801 804 805 808   
## 0.216060379 0.023759110 0.195807331 0.094810876 0.011416233 0.190594131   
## 809 810 812 813 815 818   
## 0.179675551 0.108223938 0.082146230 0.100573416 0.016327425 0.062724458   
## 823 825 826 830 833 843   
## 0.098370798 0.098923200 0.264585555 0.389681247 0.091324688 0.405172081   
## 851 853 854 858 863 864   
## 0.212843204 0.101366519 0.383607380 0.254550758 0.114034483 0.108750947   
## 867 868 869 874 880 887   
## 0.102596820 0.025142611 0.301961176 0.106314061 0.068766462 0.381568727   
## 889 890 892 894 897 906   
## 0.130902425 0.607452300 0.050854281 0.100335607 0.210353538 0.028940393   
## 907 908 909 910 916 918   
## 0.177659128 0.015460602 0.232412735 0.584258198 0.237291791 0.419483661   
## 921 922 933 934 935 939   
## 0.151865250 0.401048486 0.385068025 0.245608877 0.223023438 0.066442542   
## 940 941 948 951 958 962   
## 0.097446105 0.211913782 0.125870251 0.192013314 0.111697020 0.262772606   
## 965 967 969 972 974 976   
## 0.162913673 0.112914983 0.126912052 0.030263189 0.055597618 0.050224221   
## 978 982 992 996 997 1001   
## 0.192758257 0.354629968 0.381156786 0.165499051 0.301692232 0.095786935   
## 1002 1010 1012 1015 1027 1032   
## 0.115747519 0.042227656 0.245077335 0.058483590 0.061109403 0.026943689   
## 1035 1036 1040 1043 1046 1049   
## 0.249693371 0.198742303 0.137036422 0.073507162 0.091223512 0.130985858   
## 1050 1059 1061 1062 1066 1068   
## 0.147700619 0.410624426 0.644326085 0.294398665 0.135124053 0.212777848   
## 1069 1071 1073 1074 1082 1091   
## 0.256059155 0.214122459 0.177827170 0.217676059 0.094887784 0.143616069   
## 1092 1096 1099 1101 1102 1103   
## 0.202178525 0.363989015 0.214880499 0.187791833 0.120165711 0.253324740   
## 1104 1105 1106 1108 1110 1111   
## 0.053713993 0.175422316 0.168585193 0.079433533 0.171120150 0.486731274   
## 1122 1127 1130 1143 1157 1160   
## 0.074637605 0.051903651 0.015196191 0.258450350 0.041784517 0.067022143   
## 1161 1163 1164 1166 1168 1173   
## 0.033729037 0.091710001 0.129627991 0.120519193 0.271194921 0.122609982   
## 1174 1177 1181 1184 1191 1194   
## 0.114469286 0.034247783 0.157106040 0.122066246 0.111203986 0.126052829   
## 1196 1197 1198 1201 1203 1210   
## 0.015325909 0.269374933 0.363323125 0.179477220 0.103473188 0.012179291   
## 1211 1212 1213 1215 1220 1224   
## 0.194488961 0.039223033 0.110066707 0.103045560 0.156646691 0.017664561   
## 1229 1231 1234 1235 1236 1238   
## 0.082486710 0.388837042 0.123214959 0.069877383 0.075111797 0.119620443   
## 1240 1245 1251 1252 1254 1258   
## 0.122999780 0.172219462 0.302126398 0.178431643 0.104516199 0.110567977   
## 1259 1272 1281 1282 1284 1286   
## 0.258673974 0.283435809 0.110510224 0.287225875 0.113875461 0.098467269   
## 1287 1288 1290 1295 1307 1309   
## 0.140000345 0.142079018 0.107947918 0.081657574 0.132025057 0.182297011   
## 1315 1324 1327 1328 1333 1336   
## 0.049118057 0.113161226 0.282025006 0.030557212 0.467443579 0.090933138   
## 1340 1346 1350 1351 1354 1355   
## 0.594160431 0.414916083 0.426494174 0.100782473 0.557182218 0.053195101   
## 1357 1360 1368 1370 1373 1376   
## 0.100516269 0.182615246 0.197477562 0.319246197 0.124399922 0.372560980   
## 1378 1382 1384 1385 1391 1392   
## 0.008623114 0.198774175 0.152062892 0.034176205 0.159850743 0.163869804   
## 1394 1399 1407 1410 1411 1413   
## 0.109760278 0.087881586 0.240859236 0.066807108 0.073457017 0.134597678   
## 1415 1416 1419 1424 1426 1431   
## 0.044804009 0.129454970 0.073497209 0.178984543 0.267377853 0.032837784   
## 1432 1434 1435 1436 1438 1440   
## 0.036034159 0.400858534 0.060176575 0.181692455 0.015892273 0.043458316   
## 1443 1445 1447 1453 1454 1455   
## 0.286475023 0.042142997 0.129306445 0.035076913 0.064820544 0.081374918   
## 1464 1467 1468   
## 0.116863428 0.081457845 0.394797528

# Convert probabilities to class (0/1) with threshold 0.5  
predicted\_class <- ifelse(probabilities > 0.5, 1, 0)  
predicted\_class

## 3 7 14 15 21 22 23 27 30 33 43 47 50 53 54 57   
## 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0   
## 59 60 62 63 65 66 70 73 82 85 87 92 97 101 102 104   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 107 109 110 113 114 116 118 123 125 131 132 135 136 140 146 147   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 149 150 154 157 173 175 176 181 182 183 184 187 188 192 198 203   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 213 214 216 219 226 228 231 233 245 246 249 253 254 257 260 267   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 268 272 273 274 276 283 284 288 295 296 299 300 301 306 308 313   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 314 317 322 324 325 333 338 340 342 346 350 351 353 354 356 359   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 360 361 362 363 364 368 369 371 377 388 390 394 395 399 404 405   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 406 408 414 417 418 427 428 429 435 438 439 440 442 443 453 454   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 465 476 483 491 492 495 496 497 498 501 505 506 509 510 511 513   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 514 518 532 543 546 552 553 556 558 567 568 569 570 572 573 574   
## 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 580 582 585 587 596 604 607 609 613 614 621 624 629 636 638 641   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 644 645 652 659 661 662 669 676 684 694 695 701 704 707 716 717   
## 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0   
## 718 720 723 731 732 733 734 735 736 738 739 745 746 748 760 761   
## 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0   
## 762 763 766 768 773 777 782 783 784 795 796 800 801 804 805 808   
## 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 809 810 812 813 815 818 823 825 826 830 833 843 851 853 854 858   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 863 864 867 868 869 874 880 887 889 890 892 894 897 906 907 908   
## 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0   
## 909 910 916 918 921 922 933 934 935 939 940 941 948 951 958 962   
## 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 965 967 969 972 974 976 978 982 992 996 997 1001 1002 1010 1012 1015   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 1027 1032 1035 1036 1040 1043 1046 1049 1050 1059 1061 1062 1066 1068 1069 1071   
## 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0   
## 1073 1074 1082 1091 1092 1096 1099 1101 1102 1103 1104 1105 1106 1108 1110 1111   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 1122 1127 1130 1143 1157 1160 1161 1163 1164 1166 1168 1173 1174 1177 1181 1184   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 1191 1194 1196 1197 1198 1201 1203 1210 1211 1212 1213 1215 1220 1224 1229 1231   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 1234 1235 1236 1238 1240 1245 1251 1252 1254 1258 1259 1272 1281 1282 1284 1286   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 1287 1288 1290 1295 1307 1309 1315 1324 1327 1328 1333 1336 1340 1346 1350 1351   
## 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0   
## 1354 1355 1357 1360 1368 1370 1373 1376 1378 1382 1384 1385 1391 1392 1394 1399   
## 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 1407 1410 1411 1413 1415 1416 1419 1424 1426 1431 1432 1434 1435 1436 1438 1440   
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0   
## 1443 1445 1447 1453 1454 1455 1464 1467 1468   
## 0 0 0 0 0 0 0 0 0

# Confusion matrix  
conf\_matrix <- table(Predicted=predicted\_class, Actual=test\_data$Attrition)  
print(conf\_matrix)

## Actual  
## Predicted 0 1  
## 0 364 66  
## 1 2 9

# Accuracy  
accuracy <- sum(diag(conf\_matrix)) / sum(conf\_matrix)  
accuracy

## [1] 0.845805

print(paste("Accuracy:", round(accuracy, 3)))

## [1] "Accuracy: 0.846"

# ROC Curve & AUC  
install.packages("pROC")

## Warning: package 'pROC' is in use and will not be installed

library(pROC)  
roc\_obj <- roc(test\_data$Attrition, probabilities)

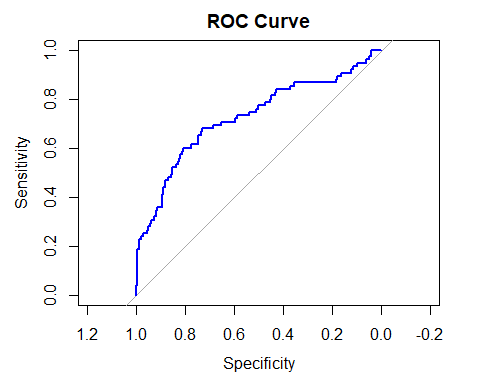
## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

roc\_obj

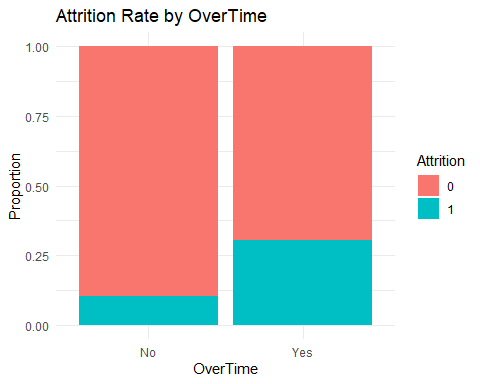
##   
## Call:  
## roc.default(response = test\_data$Attrition, predictor = probabilities)  
##   
## Data: probabilities in 366 controls (test\_data$Attrition 0) < 75 cases (test\_data$Attrition 1).  
## Area under the curve: 0.7255

plot(roc\_obj, col="blue", main="ROC Curve")

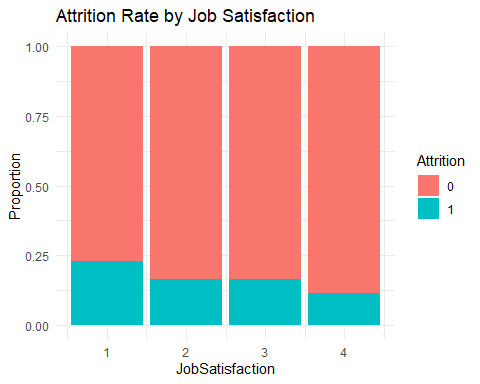


#auc\_value <- auc(roc\_obj)  
#print(paste("AUC:", round(auc\_value, 3)))

# Attrition by OverTime  
ggplot(data, aes(x=OverTime, fill=factor(Attrition))) +  
 geom\_bar(position="fill") +  
 labs(title="Attrition Rate by OverTime", y="Proportion", fill="Attrition") +  
 theme\_minimal()



# Attrition by JobSatisfaction  
ggplot(data, aes(x=JobSatisfaction, fill=factor(Attrition))) +  
 geom\_bar(position="fill") +  
 labs(title="Attrition Rate by Job Satisfaction", y="Proportion", fill="Attrition") +  
 theme\_minimal()



Final Interpretation :

Developed a logistic regression model to predict employee attrition using HR data. Key factors affecting attrition include overtime, job satisfaction, years at company, age, and monthly income. The model achieved an accuracy of 84.5% , helping HR departments identify at-risk employees and implement proactive retention strategies.