model3\_dt

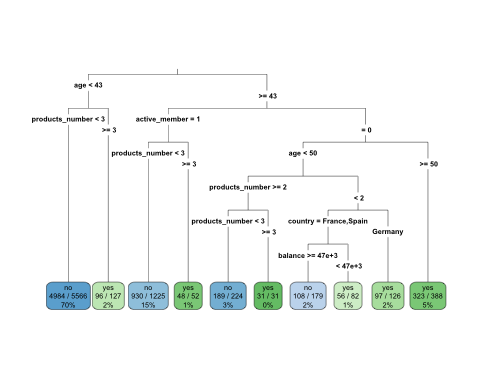
2024-12-08

# Load libraries  
library(rpart)  
library(rpart.plot)  
library(caret)

## Loading required package: ggplot2

## Loading required package: lattice

data <- read.csv("/Users/ahzenthu/Downloads/Bank Customer Churn Prediction.csv")  
  
data$country <- factor(data$country)  
data$gender <- factor(data$gender)  
data$churn <- factor(ifelse(as.numeric(data$churn) < 0.2, 'no', 'yes'))  
  
set.seed(123) # For reproducibility  
sample\_size <- floor(0.8 \* nrow(data))  
train\_ind <- sample(seq\_len(nrow(data)), size = sample\_size)  
train <- data[train\_ind, ]  
test <- data[-train\_ind, ]  
  
dt\_model <- rpart(  
 churn ~ credit\_score + country + gender + age + tenure + balance +  
 products\_number + credit\_card + active\_member + estimated\_salary,  
 data = train,  
 method = "class"  
)  
rpart.plot(dt\_model, type = 3, extra = 102, fallen.leaves = TRUE)



dt\_pred <- predict(dt\_model, newdata = test, type = "class")  
confusion <- confusionMatrix(dt\_pred, test$churn)  
print(confusion)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction no yes  
## no 1550 231  
## yes 47 172  
##   
## Accuracy : 0.861   
## 95% CI : (0.8451, 0.8759)  
## No Information Rate : 0.7985   
## P-Value [Acc > NIR] : 2.095e-13   
##   
## Kappa : 0.4792   
##   
## Mcnemar's Test P-Value : < 2.2e-16   
##   
## Sensitivity : 0.9706   
## Specificity : 0.4268   
## Pos Pred Value : 0.8703   
## Neg Pred Value : 0.7854   
## Prevalence : 0.7985   
## Detection Rate : 0.7750   
## Detection Prevalence : 0.8905   
## Balanced Accuracy : 0.6987   
##   
## 'Positive' Class : no   
##

The model is very effective at identifying customers who will not churn (97.06% sensitivity). This is critical when predicting outcomes for the majority class. The overall accuracy of 86.1%, indicating solid overall performance. Kappa value of 0.4792 indicates moderate agreement between predictions and actual outcomes.

Specificity is only 42.68%, meaning the model struggles to correctly identify customers who will churn. The majority class dominates (not churn), which may bias the model towards predicting “no churn.” overall, The decision tree model performs well in predicting customers who will not churn but struggles with identifying churn cases.