MV-lab-7.R

saiup

2022-12-01

library(caret)

## Loading required package: ggplot2

## Loading required package: lattice

library(tidyverse)

## ── Attaching packages  
## ───────────────────────────────────────  
## tidyverse 1.3.2 ──

## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.4.1   
## ✔ readr 2.1.2 ✔ forcats 0.5.2   
## ✔ purrr 0.3.4   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ purrr::lift() masks caret::lift()

library(rsample)  
library(rpart.plot)

## Loading required package: rpart

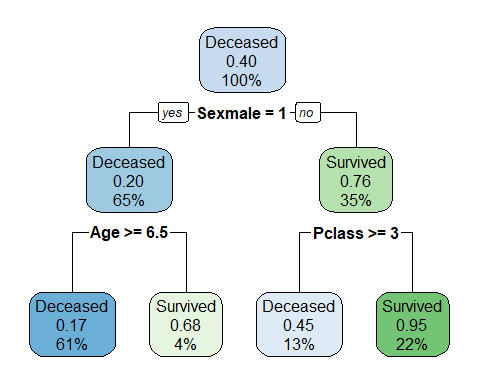
titanic <- read\_csv("titanic.csv")

## New names:  
## Rows: 499 Columns: 9  
## ── Column specification  
## ──────────────────────────────────────────────────────── Delimiter: "," chr  
## (2): Name, Sex dbl (7): ...1, Survived, Pclass, Age, SibSp, Parch, Fare  
## ℹ Use `spec()` to retrieve the full column specification for this data. ℹ  
## Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## • `` -> `...1`

titanic\_hidden <- read\_csv("titanic\_hidden.csv")

## New names:  
## Rows: 215 Columns: 9  
## ── Column specification  
## ──────────────────────────────────────────────────────── Delimiter: "," chr  
## (2): Name, Sex dbl (7): ...1, Survived, Pclass, Age, SibSp, Parch, Fare  
## ℹ Use `spec()` to retrieve the full column specification for this data. ℹ  
## Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## • `` -> `...1`

titanic <- titanic %>% select(-Name)  
titanic\_hidden <- titanic\_hidden %>% select(-Name)  
titanic$Survived <- factor(titanic$Survived, labels = c("Deceased","Survived"))  
titanic\_hidden$Survived <- factor(titanic\_hidden$Survived, labels =c("Deceased", "Survived"))  
  
model\_DTC <- train(Survived ~ .,data = titanic, method = "rpart")  
rpart.plot(model\_DTC$finalModel)



log\_prediction <- predict(model\_DTC, titanic\_hidden)  
confusionMatrix(log\_prediction, titanic\_hidden$Survived)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction Deceased Survived  
## Deceased 117 44  
## Survived 6 48  
##   
## Accuracy : 0.7674   
## 95% CI : (0.7052, 0.8222)  
## No Information Rate : 0.5721   
## P-Value [Acc > NIR] : 1.650e-09   
##   
## Kappa : 0.4989   
##   
## Mcnemar's Test P-Value : 1.672e-07   
##   
## Sensitivity : 0.9512   
## Specificity : 0.5217   
## Pos Pred Value : 0.7267   
## Neg Pred Value : 0.8889   
## Prevalence : 0.5721   
## Detection Rate : 0.5442   
## Detection Prevalence : 0.7488   
## Balanced Accuracy : 0.7365   
##   
## 'Positive' Class : Deceased   
##

model\_random<- train(Survived ~ .,data = titanic,method = 'ranger',num.trees = 200,importance = 'impurity')  
  
log\_prediction2 <- predict(model\_random, titanic\_hidden)  
confusionMatrix(log\_prediction2, titanic\_hidden$Survived)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction Deceased Survived  
## Deceased 113 30  
## Survived 10 62  
##   
## Accuracy : 0.814   
## 95% CI : (0.7554, 0.8636)  
## No Information Rate : 0.5721   
## P-Value [Acc > NIR] : 4.671e-14   
##   
## Kappa : 0.6093   
##   
## Mcnemar's Test P-Value : 0.002663   
##   
## Sensitivity : 0.9187   
## Specificity : 0.6739   
## Pos Pred Value : 0.7902   
## Neg Pred Value : 0.8611   
## Prevalence : 0.5721   
## Detection Rate : 0.5256   
## Detection Prevalence : 0.6651   
## Balanced Accuracy : 0.7963   
##   
## 'Positive' Class : Deceased   
##