Hepatitis_Logistic-Regression.R

rocka

2023-11-13

```r
# Loading 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
```

```r
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.3.2
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

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

```r
data <- read.csv("C:/Users/rocka/OneDrive/Documents/cleaned_hepatitis.csv", as.is=FALSE)


# Converting Class (Live or Die) to a factor
data$Class <- as.factor(data$Class)

# Splitting the data into training and testing sets
set.seed(123)
training.samples <- createDataPartition(data$Class, p = 0.75, list = FALSE)
train.data  <- data[training.samples, ]
test.data <- data[-training.samples, ]

# Training the model
model <- glm(Class ~ ., family = binomial(), data = train.data)
```

```
## Warning: glm.fit: algorithm did not converge
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
# Make predictions using the model
predictions <- predict(model, newdata = test.data, type = "response")

# Converting probabilities to class labels
predictions <- ifelse(predictions > 0.5, "Live", "Die")

predictions <- factor(predictions, levels = levels(test.data$Class))

# Evaluating the model
confusionMatrix(predictions, test.data$Class)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction Die Live
##       Die   3    2
##       Live  0   14
##
##              Accuracy : 0.8947
##                95% CI : (0.6686, 0.987)
##   No Information Rate : 0.8421
##   P-Value [Acc > NIR] : 0.4038
##
##                 Kappa : 0.6885
##
##  Mcnemar's Test P-Value : 0.4795
##
##           Sensitivity : 1.0000
##           Specificity : 0.8750
##        Pos Pred Value : 0.6000
##        Neg Pred Value : 1.0000
##            Prevalence : 0.1579
##        Detection Rate : 0.1579
##  Detection Prevalence : 0.2632
##     Balanced Accuracy : 0.9375
##
##      'Positive' Class : Die
##
```