

Laboratorium Analizy Procesów Uczenia.

Data wykonania ćwiczenia:

26.04.2024

Rok studiów:

1

Semestr:

1

Grupa studencka:

1b

Grupa laboratoryjna:

-

Ćwiczenie nr

3.1

Temat: Modelowanie procesów uczenia maszynowego w pakiecie mlr.
Trenowanie, ocena i porównywanie modeli w pakiecie mlr.

Osoby wykonujące ćwiczenia:

1. Gracjan Wackermann

Katedra Informatyki i Automatyki

1. Cel ćwiczenia:

Celem ćwiczenia było uczenie maszynowe za pomocą pakietu mlr.

2. Zadanie do wykonania:

Zadanie dotyczy konstruowania drzew decyzyjnych oraz reguł klasyfikacyjnych na podstawie zbioru danych (library(MASS lub datasets)).

3. mtcars

- Wariant nr. 3 -

Uzyskany kod:

```
rm(list = ls())

# Instalacja potrzebnych pakietów, jeśli nie są już
zainstalowane
install.packages(c("mlr", "C50", "rpart", "party", "readr",
                  "dplyr", "magrittr"))

# Załadowanie potrzebnych bibliotek
library(mlr)
library(C50)
library(rpart)
library(party)
library(readr)
library(dplyr)
library(magrittr)
library(rpart.plot)

# Załadowanie zestawu danych mtcars
data("mtcars")

# Utworzenie nowej zmiennej binarnej 'fast' na podstawie
mediany mpg
mtcars$fast <- ifelse(mtcars$mpg >= median(mtcars$mpg), 1, 0)
mtcars$fast <- factor(mtcars$fast, levels = c(0, 1), labels =
c("slow", "fast"))
```

```

# Ustawienie ziarna dla reprodukowalności wyników
set.seed(123)
trainIndex <- sample(1:nrow(mtcars), 0.7 * nrow(mtcars))
trainData <- mtcars[trainIndex, ]
testData <- mtcars[-trainIndex, ]

# Definicja zadania klasyfikacji
task <- makeClassifTask(data = trainData, target = "fast")

# Trenowanie modelu drzewa decyzyjnego C5.0
learner <- makeLearner("classif.C50", predict.type =
"response")
model <- train(learner, task)

# Predykcja na zestawie testowym
pred <- predict(model, newdata = testData)

# Ocena modelu
performance <- performance(pred, measures = list(acc, mmce))
print(performance)

# Definicja uczących
learners <- list(
  makeLearner("classif.rpart"),
  makeLearner("classif.C50"),
  makeLearner("classif.randomForest")
)

# Benchmarking
resampling <- makeResampleDesc("CV", iters = 5)
benchmark_result <- benchmark(learners, tasks = task,
resampling = resampling)
print(benchmark_result)

# Trenowanie modelu rpart
rpart_model <- rpart(fast ~ ., data = trainData)
rpart.plot(rpart_model)

# Zapisanie wytrenowanego modelu
save(model, file = "decision_tree_model.RData")

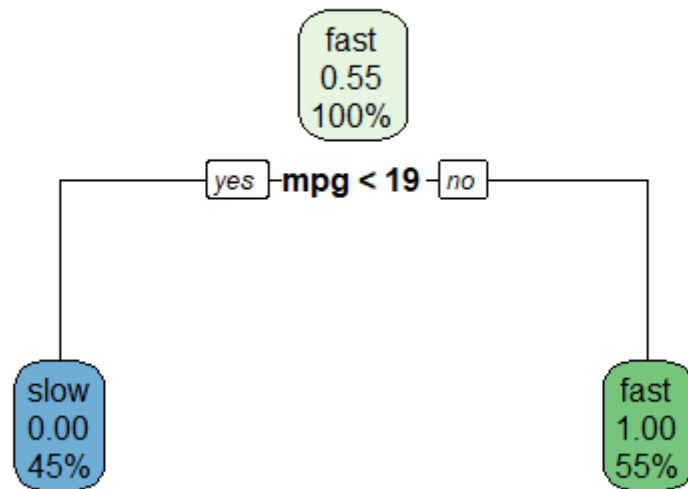
# Zapisanie wyników benchmarkingu
save(benchmark_result, file = "benchmark_results.RData")

# Zapisanie wyników oceny
write.csv(as.data.frame(performance),
"performance_results.csv")

```

3. Wnioski:

- Ten kod R służy do budowania i oceny modeli klasyfikacji na podstawie zbioru danych mtcars. Zbiór danych mtcars zawiera dane o różnych modelach samochodów, w tym ich zużycie paliwa (mpg) i inne cechy. W tym przypadku, tworzony jest nowy atrybut binarny fast, który jest określany na podstawie mediany zużycia paliwa (mpg). Samochody z zużyciem paliwa powyżej mediany są klasyfikowane jako fast, a pozostałe jako slow. Jednakże z niewiadomych przyczyn, na grafie wskazywany jest tylko i wyłącznie jeden obiekt mimo zadeklarowania całej bazy danych.



Rysunek 1 - Wynik działania.