

SPRAWOZDANIE

Zajęcia: Analiza Procesów Ucznia

Prowadzący: prof. dr hab. Vasyl Martsenyuk

Laboratorium 4

Data 28.04.2023

Temat: Użycie drzewa decyzji

Wariant 1

Rafał Klinowski
Informatyka II stopień,
Stacjonarne,
1 semestr,
Gr. a

1. Polecenie: Wariant 1

Zadanie dotyczy prognozowania oceny klientów (w skali 5-punktowej, $\text{Error} < 5\%$) urządzeń RTV AGD. Używając metody indukcji drzewa decyzji C5.0 opracować plik w języku R z wykorzystaniem paczki *C50*. Sprawozdania w postaci pliku R, obrazu drzewa decyzji oraz wyników z konsoli

1. Smartfon Samsung z systemem Android 6, ośmiordziennej procesorem. Uwzględniamy następujące dane: wyświetlacz, pamięć RAM, pamięć wbudowana, aparat foto. Komunikacja (wifi itp) nie ma znaczenia. Dane (>10 smartfonów) pobrać ze strony <http://www.euro.com.pl>

2. Wprowadzane dane:

Dane do ćwiczenia zostały zaczerpnięte z pierwszego z laboratoriów (plik CSV zawierający opis 15 smartfonów).

3. Wykorzystane komendy:

Poniżej można znaleźć wszystkie wykorzystane komendy:

Autor: Rafal Klinowski, wariant: 1.

```
setwd('C:\\Users\\klino\\Pulpit\\Studia magisterskie\\APU\\Lab4')
```

```
# Załadowanie danych o smartfonach (z lab1).  
smartfony <- read.csv('smartfony.csv')
```

```
# Instalacja pakietu C50.  
install.packages("C50")  
library(C50)
```

```
# Konwersja wyjścia z drzewa (ocen klientów) na typ factor.  
smartfony$oceny_klientow <- as.factor(smartfony$oceny_klientow)
```

```
# Stworzenie drzewa - prognozujemy oceny na podstawie innych wartości.  
tree <- C5.0(smartfony[, c('pamiec_ram', 'pamiec_wbudowana', 'aparat_foto', 'cena')],  
             smartfony[, 'oceny_klientow'])
```

```
# Wizualizacja drzewa.  
plot(tree)
```

```
# Dane o drzewie.  
summary(tree)
```

```
# -----
# Powtórzenie powyższego z podziałem na zbior treningowy i testowy.
# W celu ewaluacji drzewa.

smartfony_train <- smartfony[1:12,]
smartfony_test <- smartfony[13:15,]

vars <- c('pamiec_ram', 'pamiec_wbudowana', 'aparat_foto', 'cena')

tree <- C5.0(x=smartfony_train[, vars],
             y=smartfony_train$oceny_klientow)

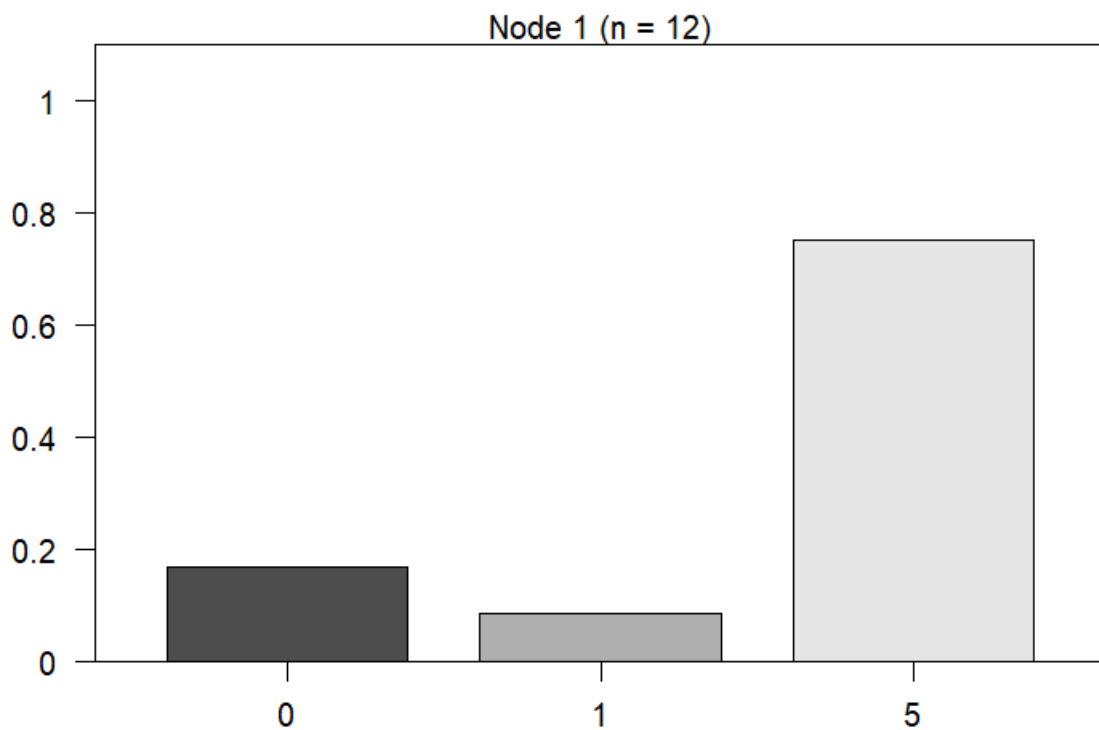
summary(tree)

plot(tree)

# Predykcja dla danych testowych.
predict(tree, newdata = smartfony_test[, vars], type="prob")
```

4. Wynik działania:

Wyniki poleceń w konsoli można znaleźć w pliku „wyniki z konsoli.txt”, link do repozytorium poniżej.



Rysunek 1. Wykres prezentujący uzyskane wyniki.

Link do repozytorium: https://github.com/Stukeley/APU_Lab4

5. Wnioski:

Dla naszych danych testowych nie udało się uzyskać efektywnego drzewa – uzyskano drzewo zawierające zaledwie jeden poziom, co można interpretować jako słabą zależność między zmiennymi wejściowymi. Podobny wynik uzyskano po usunięciu ze zbioru wejściowego kolumny ‘cena’.

Prawdopodobnie wykorzystana ilość smartfonów była zbyt niska, aby uzyskać satysfakcjonujące wyniki.