**SPRAWOZDANIE**

Zajęcia: Analiza Procesów Uczenia

Prowadzący: prof. dr hab. Vasyl Martsenyuk

**Laboratorium 4**

Data 28.04.2023

**Temat:**​ Użycie drzewa decyzji

**Wariant 1**

Rafał Klinowski

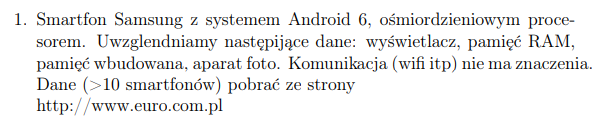
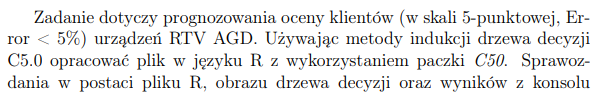
Informatyka II stopień,

Stacjonarne,

1 semestr,

Gr. a

1. **Polecenie: Wariant 1**



1. **Wprowadzane dane:**

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

1. **Wykorzystane komendy:**

Poniżej można znaleźć wszystkie wykorzystane komendy:

# Autor: Rafal Klinowski, wariant: 1.

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

# Zaladowanie danych o smartfonach (z lab1).

smartfony <- read.csv('smartfony.csv')

# Instalacja pakietu C50.

install.packages("C50")

library(C50)

# Konwersja wyjscia z drzewa (ocen klientow) na typ factor.

smartfony$oceny\_klientow <- as.factor(smartfony$oceny\_klientow)

# Stworzenie drzewa - prognozujemy oceny na podstawie innych wartosci.

tree <- C5.0(smartfony[, c('pamiec\_ram', 'pamiec\_wbudowana', 'aparat\_foto', 'cena')],

smartfony[, 'oceny\_klientow'])

# Wizualizacja drzewa.

plot(tree)

# Dane o drzewie.

summary(tree)

# -----------------------------

# Powtorzenie powyzszego z podzialem 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)

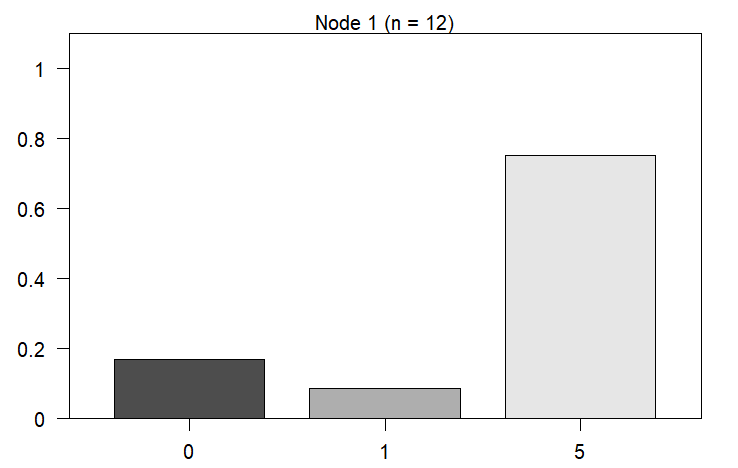
plot(tree)

# 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>

1. **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.