

# **SPRAWOZDANIE**

Zajęcia: Analiza Procesów Ucznia

Prowadzący: prof. dr hab. Vasyl Martsenyuk

**Laboratorium: 4**

**Data: 25.05.2020**

**Temat: „Uczenie maszynowe z użyciem drzew decyzyjnych”**

**Wariant: 1**

Jacek Adamczyk

Informatyka II stopień,

Stacjonarne,

1 semestr,

Gr. A

<https://github.com/jacekaGIT/ATH-1g>

## 1. Polecenia dla wariantu 1:

Celem ćwiczenia jest podejmowanie decyzji za pomocą indukcji drzewa decyzji na podstawie algorytmu C5.0

Polecenie:

Zadanie dotyczy prognozowania oceny klientów (w skali 5-punktowej, 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 (dowolny plik tekstowy) zachować w zdalnym repozytorium (np Github) link na który wysłać w mailu z tematem **SMPD\_7\_Gr\_numer\_grupy** na adres mailowy **vmartsenyuk@ath.bielsko.pl**

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. Skrypt:

Po uruchomieniu skryptu zostaną wykonane kolejno wszystkie polecenia zadania. Instalacja pakietu „C5.0” została wyłączona (ustawiona jako komentarz) żeby niepotrzebnie nie instalować pakietu przy każdym uruchomieniu skryptu.

```
#..... Jacek Adamczyk, sem 1, II st .....
#..... Lab 4, gr A, wariant 1 .....
#Uczenie maszynowe z użyciem drzew decyzyjnych
#.....

#install.packages("C50")
library("C50")

setwd("C:/Users/jacek/Dysk Google/ATH 1/APU/Lab4")
smartfony <- read.csv("JacekAdamczyk_APU_Lab1.csv")
smartfony
View(smartfony)
summary(smartfony)

smartfony$ocena <- factor(smartfony$ocena)
smartfony

drzewo <- C5.0.default(x=smartfony[,-9], y=smartfony$ocena )

summary(drzewo)
drzewo
plot(drzewo)

predict.C5.0(drzewo, smartfony[,-9])
predict.C5.0(drzewo, smartfony[,-9], type = "class")
```

### 3. Wyniki działania:

Poniżej przedstawiono wyniki działania skryptu. Pełne wydruki z konsoli można znaleźć w repozytorium GitHub (adres na stronie tytułowej). Parametry smartfonów wykorzystane jako dane uczące dla drugiej części zadania są zawarte w pliku „JacekAdamczyk\_APU\_Lab1.csv”.

smartfony										
	X	nazwy	wyswietlacz	pamiec_RAM	pamiec_wbudowana	aparat_foto	cena	liczba_ocen	ocena	status_opinii
1	1	Galaxy A Quantum	6.70	8	128	64	2268	7	5	mniej niż 50 opinii
2	2	Galaxy A21S	6.50	4	64	48	1070	4	4	mniej niż 50 opinii
3	3	Galaxy A41	6.10	4	64	48	1299	6	4	mniej niż 50 opinii
4	4	Galaxy S20	6.70	12	128	64	5999	3	3.5	mniej niż 50 opinii
5	5	Galaxy A71	6.70	6	128	64	1528	3	3	mniej niż 50 opinii
6	6	Galaxy M21	6.40	6	64	48	999	47	3.5	mniej niż 50 opinii
7	7	Galaxy Z Flip	6.70	8	256	12	6600	17	4	mniej niż 50 opinii
8	8	Galaxy S10 Lite	6.70	8	128	48	2799	41	5	mniej niż 50 opinii
9	9	Galaxy A51	6.50	4	128	48	1699	19	4.5	mniej niż 50 opinii
10	10	Galaxy Note10	6.75	12	256	12	4999	33	5	mniej niż 50 opinii
11	11	Galaxy Xcover4S	5.00	3	32	16	1119	35	4.5	mniej niż 50 opinii
12	12	Galaxy XcoverPro	6.30	4	64	25	2199	28	5	mniej niż 50 opinii
13	13	Galaxy A10	6.20	2	32	13	699	11	4	mniej niż 50 opinii
14	14	Galaxy A80	6.70	8	128	48	2399	87	4.5	50-100 opinii
15	15	Galaxy A40	5.70	4	64	16	999	406	5	więcej niż 100 opinii

Rysunek 1. Dane smartfonów

- Wyniki:

```
> summary(drzewo)
```

Call:

```
C5.0.default(x = smartfony[, -9], y = smartfony$ocena)
```

```
C5.0 [Release 2.07 GPL Edition]    Mon May 25 08:49:05 2020
```

```
-----  
Class specified by attribute `outcome'
```

```
Read 15 cases (10 attributes) from undefined.data
```

```
Decision tree:
```

```
status_opinii = 50-100 opinii: 4.5 (1)  
status_opinii = więcej niż 100 opinii: 5 (1)  
status_opinii = mniej niż 50 opinii:  
...pamiec_RAM <= 4:  
  ...liczba_ocen <= 11: 4 (3)  
  :   liczba_ocen > 11: 4.5 (3/1)  
  pamiec_RAM > 4:  
  ...pamiec_RAM <= 6: 3 (2/1)  
  :   pamiec_RAM > 6: 5 (5/2)
```

Evaluation on training data (15 cases):

Decision Tree  
-----  
Size Errors

6 4(26.7%) <<

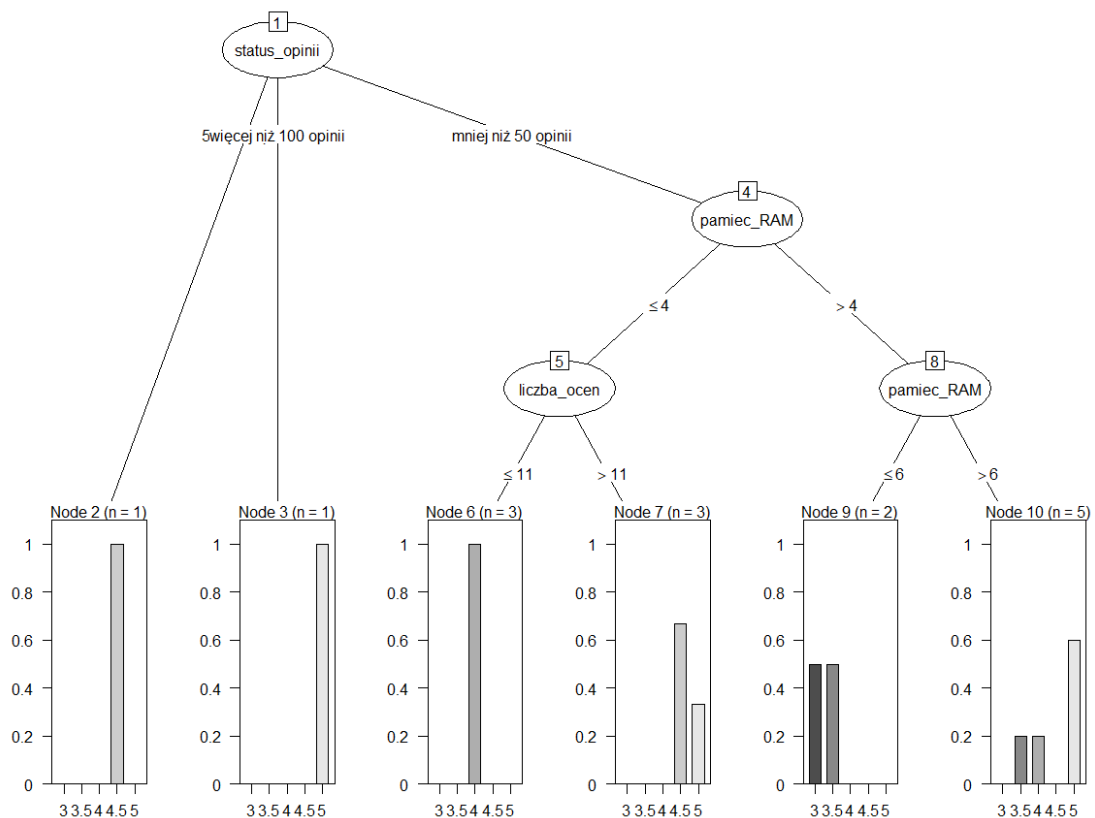
(a)	(b)	(c)	(d)	(e)	<-classified as
1					(a): class 3
1				1	(b): class 3.5
		3		1	(c): class 4
			3		(d): class 4.5
			1	4	(e): class 5

Attribute usage:

100.00%status\_opinii  
86.67%pamiec\_RAM  
40.00%liczba\_ocen

Time: 0.0 secs

- Drzewo decyzyjne:



Rysunek 2. Drzewo decyzyjne

- Wyniki predykcji:

```
> predict.C5.0(drzewo, smartfony[, -9])
[1] 5 4 4 5 3 3 5 5 4.5 5 4.5 4.5 4 4.5 5
Levels: 3 3.5 4 4.5 5
> predict.C5.0(drzewo, smartfony[, -9], type = "class")
[1] 5 4 4 5 3 3 5 5 4.5 5 4.5 4.5 4 4.5 5
Levels: 3 3.5 4 4.5 5
```

#### 4. Wnioski:

Wyniki predykcji oceny różnią się od oceny wystawionej przez klientów. Jednak nie występują bardzo duże różnice. Wynikają one z tego że klienci nie zawsze kierują się obiektywnymi danymi technicznymi a wpływ na opinie ma np. moda lub recenzje.

Nr tel.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Klienci	5	4	4	3.5	3	3.5	4	5	4.5	5	4.5	5	4	4.5	5
Predykcja	5	4	4	5	3	3	5	5	4.5	5	4.5	4.5	4	4.5	5

*Rysunek 3. Porównanie wyników predykcji z ocenami klientów*

Wpływ na wyniki predykcji może mieć również zawyżona przez producenta cena w stosunku do parametrów. Dotyczy to na pewno pozycji 4.

W badaniu wystąpił również dosyć duży błąd = 26,7%. Dużo większy niż założony w zadaniu. Wpływ na to mają bezpośrednio wartości danych dla telefonów. Aby zmniejszyć błąd należałoby je modyfikować a to nie jest dopuszczalne w takich badaniach. W efekcie uzyskalibyśmy zafałszowane wyniki.