

SPRAWOZDANIE

Zajęcia: Analiza procesów uczenia

Prowadzący: prof. dr hab. inż. Vasyl Martsenyuk

| | |
|--|--|
| Laboratorium Nr 4 Data 21.04.2023 Temat: Uczenie maszynowe z użyciem drzew decyzyjnych Wariant 4 | Maksymilian Grygiel Informatyka II stopień, stacjonarne, Semestr I, gr.1a |
|--|--|

Link do repozytorium: <https://github.com/Maksiolo20/APU>

Zadania:

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.

Wariant 4: Lodówka. Uwzględniay następujące dane: pojemność użytkową chłodziarku, pojemność użytkową zamrażarki oraz jakieś inne parametry.

Wykonanie zadania:

Instalacja oraz użycie paczki C50:

```
> install.packages("C50")
Instalowanie pakietu w 'C:/Users/MaksioloLaptop/Documents/R/win-library/4.1'
(ponieważ 'lib' nie jest określony)
Instalowanie dodatkowych zależności 'plyr', 'reshape2', 'libcoin', 'mvtnorm', 'Formula', 'inum', 'Cubist',
'partykit'

package 'plyr' successfully unpacked and MD5 sums checked
package 'reshape2' successfully unpacked and MD5 sums checked
package 'libcoin' successfully unpacked and MD5 sums checked
package 'mvtnorm' successfully unpacked and MD5 sums checked
package 'Formula' successfully unpacked and MD5 sums checked
package 'inum' successfully unpacked and MD5 sums checked
package 'Cubist' successfully unpacked and MD5 sums checked
package 'partykit' successfully unpacked and MD5 sums checked
package 'C50' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\MaksioloLaptop\AppData\Local\Temp\RtmpYJit6d\downloaded_packages
```

```
> library("C50")
Warning message:
pakiet 'C50' został zbudowany w wersji R 4.1.3
```

Ustalenie ścieżki domyślnej na folder przeznaczony do zadania 4:

```
> setwd("C:/Users/MaksioloLaptop/Desktop/mgr/APU/lab4")
```

Łaadowanie danych:

```
> dane <- read.csv("lodowki.csv")
```

Zamienienie kolumny decyzyjnej w faktor:

```
> dane$Ocena <- factor(dane$Ocena)
```

Wypisanie kilku pierwszych danych:

```
> head(dane)
```

| | Model | Pojemnosc.uzytkowa.lodowki | Pojemnosc.uzytkowa.zamrazarki | wysokosc | Poziom.halasu |
|---|-----------------------|----------------------------|-------------------------------|----------|---------------|
| 1 | LG GSL761PZU | 405 | 196 | 179 | 39 |
| 2 | Amica FD2325.4XI | 160 | 44 | 144 | 40 |
| 3 | Samsung RB37J501MB1 | 255 | 98 | 201 | 39 |
| 4 | Samsung RB33N341MSS | 217 | 98 | 185 | 38 |
| 5 | Samsung RS50N3913BC | 357 | 144 | 179 | 43 |
| 6 | LG Premium GBB59PZRZS | 225 | 75 | 190 | 37 |

```
Ocena
1      3
2      2
3     4.5
4     3.5
5      4
6     4.5
> |
```

Zbudowanie drzewa. Kolumna Ocena jest daną wyjściową:

```
> treeModel<-C5.0(x=dane[, -6], y=dane$Ocena)
```

Wypisanie drzewa:

```
> treeModel
```

Call:

```
C5.0.default(x = dane[, -6], y = dane$Ocena)
```

Classification Tree

Number of samples: 10

Number of predictors: 5

Tree size: 3

Non-standard options: attempt to group attributes

Wypisanie podsumowania drzewa:

```
> summary(treeModel)
```

```
Call:
```

```
C5.0.default(x = dane[, -6], y = dane$Ocena)
```

```
C5.0 [Release 2.07 GPL Edition]
```

```
Fri Apr 28 21:01:55 2023
```

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

```
Read 10 cases (6 attributes) from undefined.data
```

```
Decision tree:
```

```
Poziom.halasu > 39: 1 (4/3)
```

```
Poziom.halasu <= 39:
```

```
...Pojemnosc.uzytkowa.lodowki <= 255: 4.5 (3/1)
```

```
    Pojemnosc.uzytkowa.lodowki > 255: 3 (3/1)
```

```
Evaluation on training data (10 cases):
```

```
Evaluation on training data (10 cases):
```

| Decision Tree | | | | | | | | | |
|---------------|----------|-----|-----|-----|-----|-----|-----|-----------------|--|
| Size | Errors | | | | | | | | |
| 3 | 5(50.0%) | << | | | | | | | |
| (a) | (b) | (c) | (d) | (e) | (f) | (g) | (h) | <-classified as | |
| 1 | | | | | | | | (a): class 1 | |
| 1 | | | | | | | | (b): class 1.5 | |
| 1 | | | | | | | | (c): class 2 | |
| | | | 2 | | | | | (d): class 3 | |
| | | | | | | 1 | | (e): class 3.5 | |
| 1 | | | | | | | | (f): class 4 | |
| | | | | | | 2 | | (g): class 4.5 | |
| | | | 1 | | | | | (h): class 5 | |

```
Attribute usage:
```

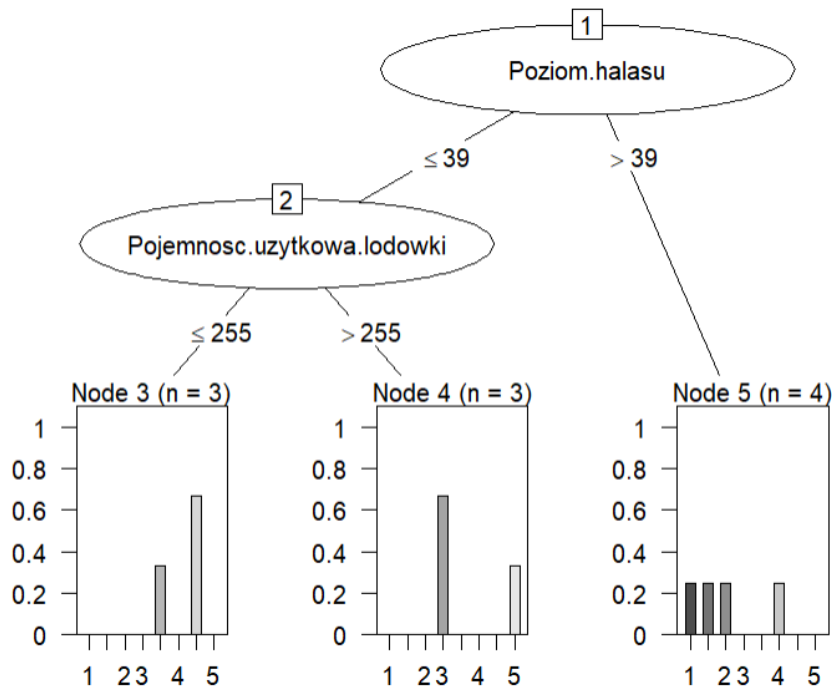
```
100.00% Poziom.halasu
```

```
60.00% Pojemnosc.uzytkowa.lodowki
```

```
Time: 0.0 secs
```

Wrysowanie drzewa do wykresu:

```
> plot(treeModel)
```



Wnioski:

Język R oraz środowisko R studio pozwalają na uczelnie tak zwaną metodą indukcji drzewa decyzji. Analizując drzewo widać, że bazową wartością jest poziom hałasu. Jeżeli jest on wyższy niż 39 to łodówki te otrzymują oceny 1, 1.5, 2, 4. Gdy jest mniejszy, czynnikiem rozdzielającym łodówki staje się pojemność użytkowa łodówki. Jeżeli jest mniejsza niż 255 to łodówki dostały oceny 3,5 (w mniejszym stopniu) oraz 4,5 (w większym stopniu). Reszta natomiast ma oceny 3 (w większym stopniu) i 5 (w większym stopniu).