

SPRAWOZDANIE

Zajęcia: Analiza Procesów Ucznienia

Prowadzący: prof. dr hab. Vasył Martsenyuk

Laboratorium 5_1

Data 28.04.2023r.

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

Wariant 5

Jarosław Waliczek
Informatyka
II stopień,
stacjonarne
1 semestr

1. Polecenie:

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

5. bacteria

2. Wprowadzane dane:

Na podstawie zbioru danych (library(MASS lub datasets))

3. Wykorzystane komendy:

Biblioteki:

```
library(MASS)
require(C50)
```

Ładowanie zbioru danych „bacteria”:

```
> data("bacteria")
>
> head(bacteria)
  y ap hilo week  ID      trt
1 y  p   hi    0 x01 placebo
2 y  p   hi    2 x01 placebo
3 y  p   hi    4 x01 placebo
4 y  p   hi   11 x01 placebo
5 y  a   hi    0 x02  drug+
6 y  a   hi    2 x02  drug+
>
> str(bacteria)
'data.frame': 220 obs. of 6 variables:
 $ y   : Factor w/ 2 levels "n","y": 2 2 2 2 2 2 1 2 2 2 ...
 $ ap  : Factor w/ 2 levels "a","p": 2 2 2 2 1 1 1 1 1 1 ...
 $ hilo: Factor w/ 2 levels "hi","lo": 1 1 1 1 1 1 1 1 2 2 ...
 $ week: int  0 2 4 11 0 2 6 11 0 2 ...
 $ ID  : Factor w/ 50 levels "x01","x02","x03",...: 1 1 1 1 2 2 2 2 3 3 ...
 $ trt : Factor w/ 3 levels "placebo","drug",...: 1 1 1 1 3 3 3 3 2 2 ...
```

Przygotowanie Factor-a:

```
> bacteria[, 'ap'] <- factor(bacteria[, 'ap'])
>
> str(bacteria)
'data.frame': 220 obs. of 6 variables:
 $ y   : Factor w/ 2 levels "n","y": 2 2 2 2 2 2 1 2 2 2 ...
 $ ap  : Factor w/ 2 levels "a","p": 2 2 2 2 1 1 1 1 1 1 ...
 $ hilo: Factor w/ 2 levels "hi","lo": 1 1 1 1 1 1 1 1 2 2 ...
 $ week: int  0 2 4 11 0 2 6 11 0 2 ...
 $ ID  : Factor w/ 50 levels "x01","x02","x03",...: 1 1 1 1 2 2 2 2 3 3 ...
 $ trt : Factor w/ 3 levels "placebo","drug",...: 1 1 1 1 3 3 3 3 2 2 ...
>
> table(bacteria$ap)
```

Przygotowanie drzewa decyzyjnego i wizualizacji:

```
> m1 <- C5.0(bacteria[1:220,-6],bacteria[1:220,6])
>
> summary(m1)
```

Call:
C5.0.default(x = bacteria[1:220, -6], y = bacteria[1:220, 6])

C5.0 [Release 2.07 GPL Edition] Fri Feb 18 00:04:31 2022

Class specified by attribute 'outcome'

Read 220 cases (6 attributes) from undefined.data

Decision tree:

```
ap = p: placebo (96)
ap = a:
...hilo = hi: drug+ (62)
      hilo = lo: drug (62)
```

Evaluation on training data (220 cases):

Decision Tree		
Size	Errors	
3	0 (0.0%)	<<

(a)	(b)	(c)	<-classified as
96			(a): class placebo
	62		(b): class drug
		62	(c): class drug+

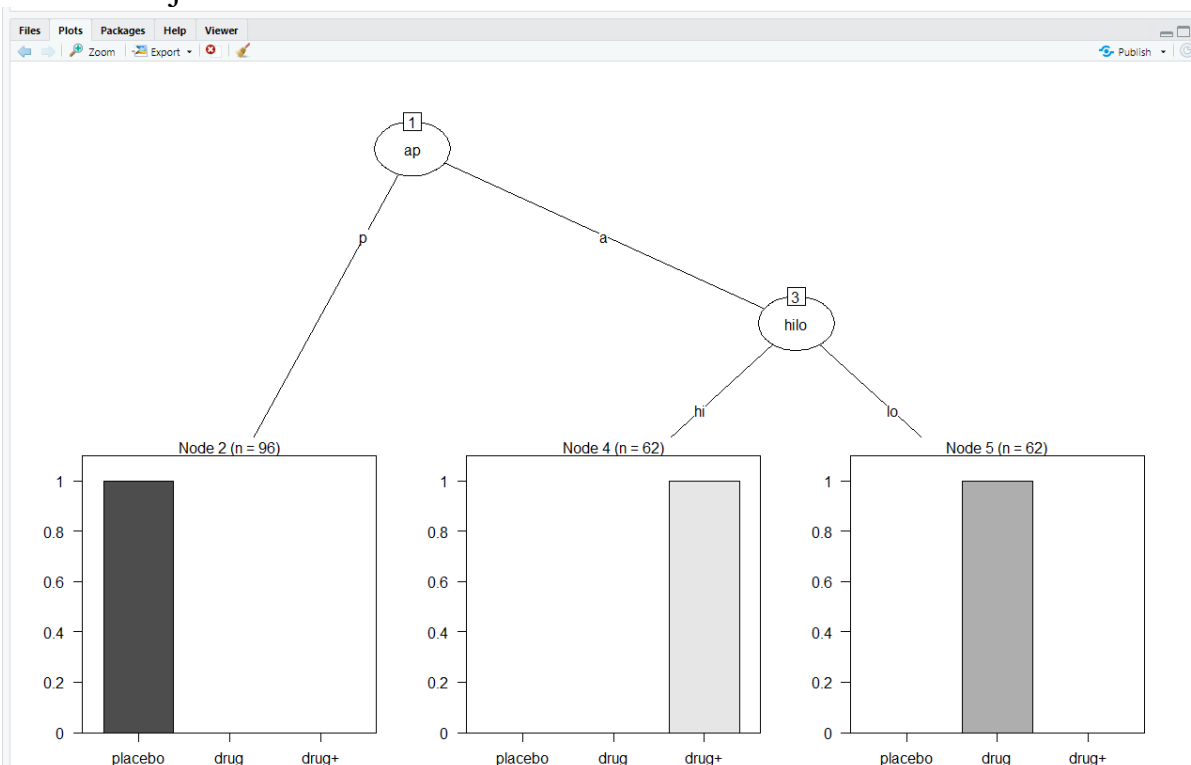
Attribute usage:

```
100.00% ap
 56.36% hilo
```

Time: 0.0 secs

```
>
> plot(m1)
```

Wizualizacja:



4. Wynik działania:

Kod programu dostępny w repozytorium: <https://github.com/Jaro233/APU.git>

Drzewo decyzyjne:

```
> m1 <- C5.0(bacteria[1:220,-6],bacteria[1:220,6])
>
> summary(m1)

Call:
C5.0.default(x = bacteria[1:220, -6], y = bacteria[1:220, 6])

C5.0 [Release 2.07 GPL Edition]          Fri Feb 18 00:04:31 2022

-----

Class specified by attribute 'outcome'

Read 220 cases (6 attributes) from undefined.data

Decision tree:

ap = p: placebo (96)
ap = a:
...hilo = hi: drug+ (62)
      hilo = lo: drug (62)

Evaluation on training data (220 cases):

      Decision Tree
      -----
      Size      Errors      <<

      3      0( 0.0%)

      (a) (b) (c) <-classified as
      --- --- ---
      96    62    62
                (a): class placebo
                (b): class drug
                (c): class drug+

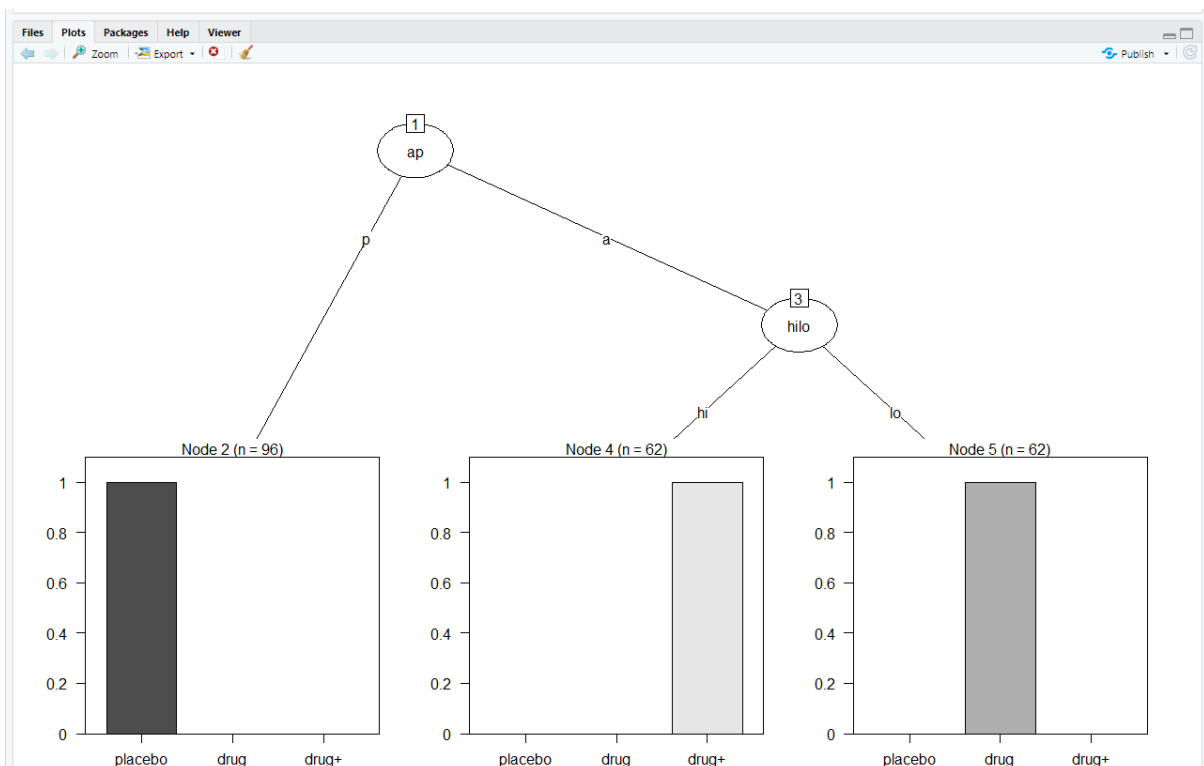
Attribute usage:

100.00% ap
56.36% hilo

Time: 0.0 secs

>
> plot(m1)
```

Wizualizacja:



5. Wnioski:

Dzięki bibliotece c50 możliwe jest uczenie maszynowe poprzez stworzenie drzewa decyzyjnego z danych wejściowych z pliku .csv, a następnie wygenerowanie wykresu na podstawie stworzonego drzewa.