

SPRAWOZDANIE

Zajęcia: Analiza Procesów Ucznienia

Prowadzący: prof. dr hab. Vasyl Martsenyuk

Laboratorium 7

Data 27.05.2019r.

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

Wariant 4

Artur Kozieł
Informatyka II stopień,
stacjonarne
1 semestr

1. Polecenie:

Celem ćwiczenia było stworzenie drzewa decyzyjnego dla zbioru danych *Aids2*.

2. Wprowadzane dane:

Użyte zostały dane *Aids2*. Są to dane dotyczące osób chorujące na AIDS. Są to dane wbudowane w standardowy pakiet R.

3. Wykorzystane komendy:

```
library("C50")           # dołączenie pakietu C.50
library("MASS")          # dołączenie pakietu Mass
data(Aids2)              # załadowanie danych Aids2
head(Aids2)              # wypisanie początkowych wartości zbioru Aids2
treeModel <- C5.0(x=Aids2[, -6], y=Aids2$T.categ) #stworzenie drzewa decyzyjnego
treeModel                # wypisanie drzewa do konsoli
summary(treeModel)       # podsumowanie uczenia
plot(treeModel)          # narysowanie drzewa w formie wykresu
```

4. Wynik działania:

Na podstawie zbioru danych *Aids2* stworzono model drzewa decyzyjnego, które pozwala przewidywać w jaki sposób dana osoba zaraziła się Aids-em.

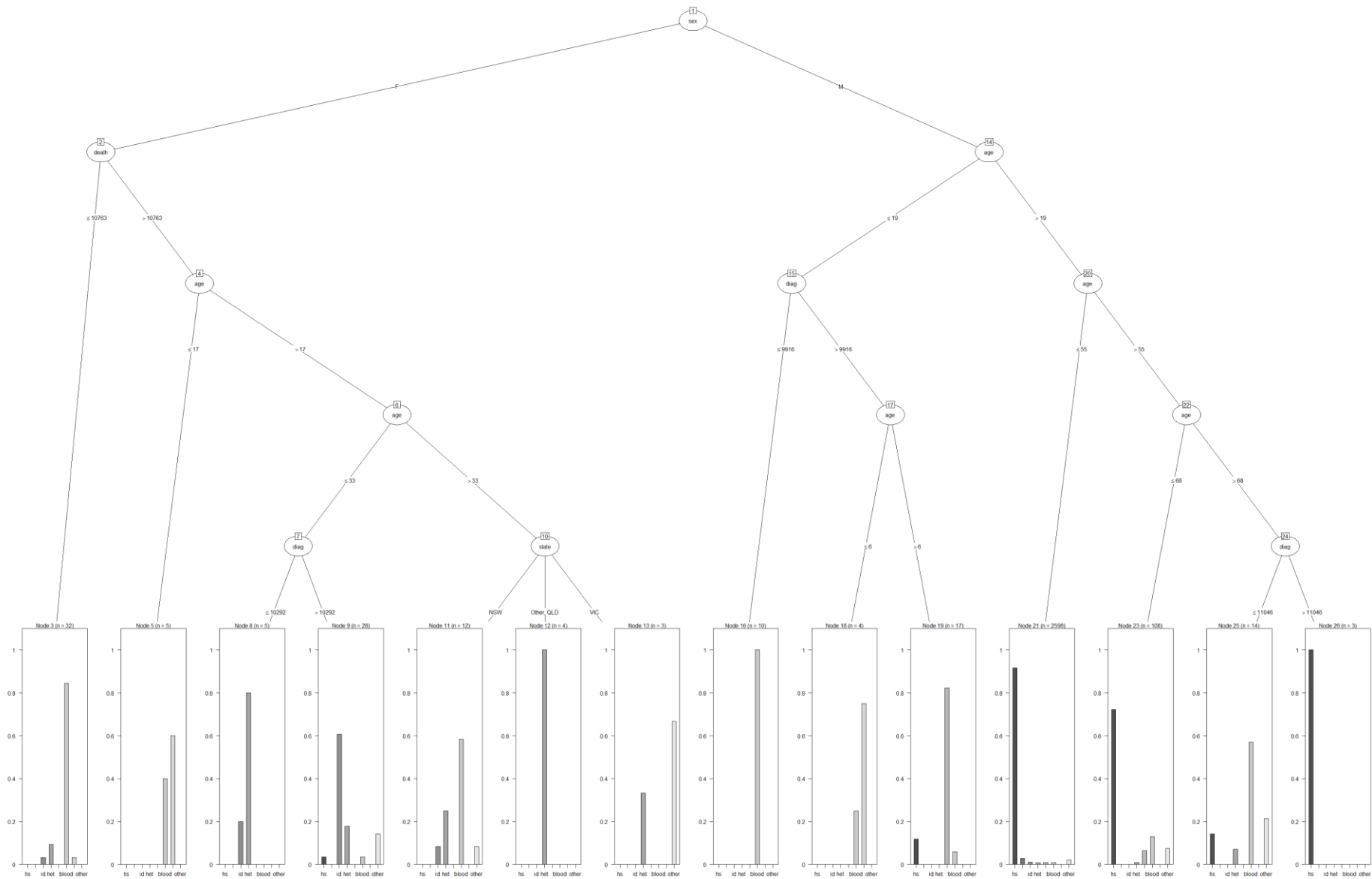
Kod programu dostępny w repozytorium: <https://github.com/arturkoziel/APU>

Kilka początkowych wierszy zbioru danych *Aids2*:

	state	sex	diag	death	status	T.categ	age
1	NSW	M	10905	11081	D	hs	35
2	NSW	M	11029	11096	D	hs	53
3	NSW	M	9551	9983	D	hs	42
4	NSW	M	9577	9654	D	haem	44
5	NSW	M	10015	10290	D	hs	39
6	NSW	M	9971	10344	D	hs	36

Jako zmienna, której wartość chcemy predykować została wybrana kolumna „*T.categ*”, czyli kategoria – sposób zarażenia człowieka.

Schemat drzewa decyzyjnego, które zostało stworzone dla danych *Aids2* wygląda następująco:



Dane są podzielone na następujące kategorie: płeć, wiek, data śmierci, data diagnozy oraz organizacja, do której należy pacjent.

Attribute usage:

100.00% sex
98.87% age
3.13% death
2.85% diag
0.67% state

5. Wnioski:

Na podstawie przeprowadzonej analizy można zauważyć, że najistotniejszą rolę w przewidywaniu sposobu zarażenia ma płeć. Dla mężczyzn kolejnym istotnym czynnikiem jest ich wiek, natomiast dla kobiet jest to data śmierci.