SPRAWOZDANIE

Zajęcia: Analiza procesów uczenia

Prowadzący: prof. dr hab. inż. Vasyl Martsenyuk

Laboratorium Nr 5 Data 28.04.2023 Temat: Modelowanie procesów uczenia maszynowego w pakecie mlr. Trenowanie, ocena i porównywanie modeli w pakecie mlr Wariant 4	Maksymilian Grygiel Informatyka II stopień, stacjonarne, Semestr I, gr.1a

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

Zadania:

Zadanie 1:

Zadanie dotyczy konstruowania drzew decyzyjnych oraz reguł klasyfikacyjnych na podstawie zbioru danych (library(MASS lub datasets)). Wariant zadania nr 4: Aids2

Wykonanie zadania:

Dołączenie pakietów c5.0 oraz Mass:

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

Załadowanie danych oraz wypisanie początkowych danych:

```
> data("Aids2")
> head(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
```

Stworzenie drzewa decyzyjnego I wypisanie go do konsoli:

```
> treeModel <- C5.0(x=Aids2[,-6],y=Aids2$T.categ)
> treeModel

Call:
C5.0.default(x = Aids2[, -6], y = Aids2$T.categ)

Classification Tree
Number of samples: 2843
Number of predictors: 6

Tree size: 14

Non-standard options: attempt to group attributes
```

Wypisanie podsumowania uczenia:

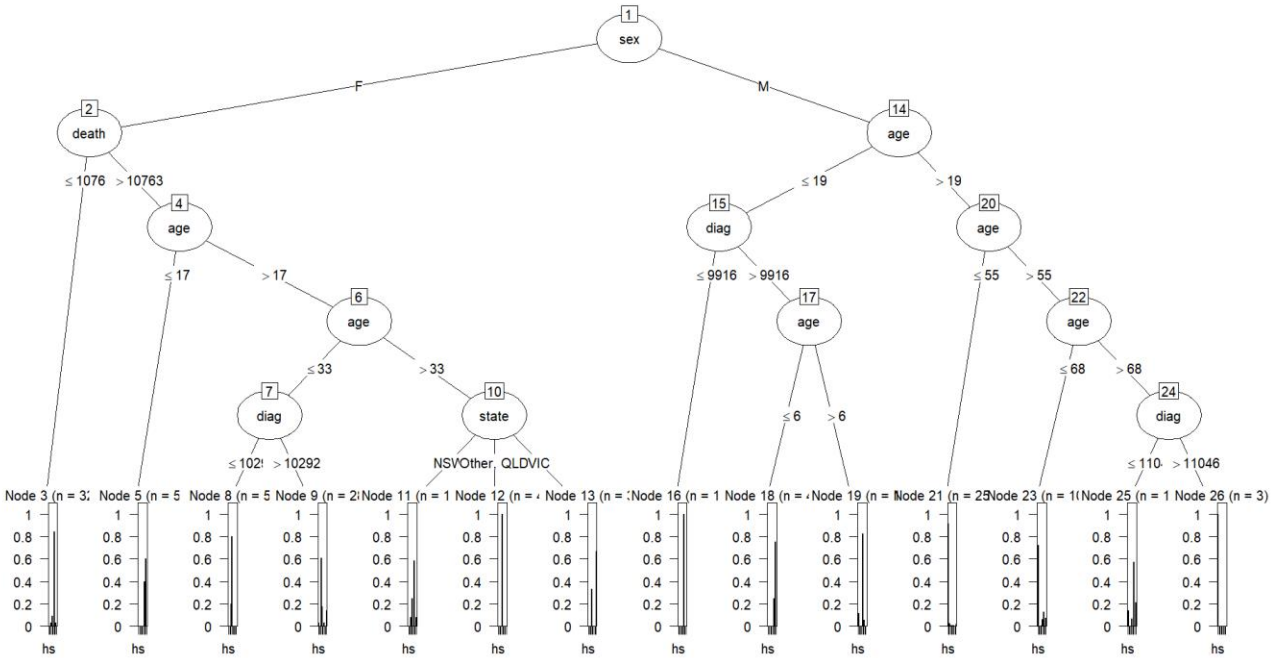
```
> summary(treeModel)

Call:
C5.0.default(x = Aids2[, -6], y = Aids2$T.categ)


C5.0 [Release 2.07 GPL Edition]         Thu May 04 00:26:40 2023
-------------------------------

Class specified by attribute `outcome'

Read 2843 cases (7 attributes) from undefined.data

Decision tree:

sex = F:
```

```
sex = F:
:...death <= 10763: blood (32/5)
:   death > 10763:
:   :...age <= 17: mother (5/2)
:       age > 17:
:       :...age <= 33:
:           :...diag <= 10292: het (5/1)
:           :   diag > 10292: id (28/11)
:           age > 33:
:           :...state = NSW: blood (12/5)
:               state in {Other,QLD}: het (4)
:               state = VIC: other (3/1)
sex = M:
:...age <= 19:
    :...diag <= 9916: blood (10)
    :   diag > 9916:
    :   :...age <= 6: mother (4/1)
    :       age > 6: haem (17/3)
    age > 19:
    :...age <= 55: hs (2598/219)
```

```
    :...age <= 55: hs (2598/219)
        age > 55:
        :...age <= 68: hs (108/30)
            age > 68:
            :...diag <= 11046: blood (14/6)
                diag > 11046: hs (3)


Evaluation on training data (2843 cases):

          Decision Tree
        ----------------
        Size      Errors

          14  284(10.0%)   <<


         (a)   (b)   (c)   (d)   (e)   (f)   (g)   (h)    <-classified as
        ----  ----  ----  ----  ----  ----  ----  ----
        2460           1           2     2                (a): class hs
          72                                              (b): class hsid
          28          17     1           2                (c): class id
          20           5     8           7           1    (d): class het
          32                       14                     (e): class haem
          37           1           1    52     3          (f): class blood
                                           1     6          (g): class mother
          60           4                 4           2    (h): class other


        Attribute usage:

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

Ostatnim krokiem jest narysowanie drzewa w formie wykresu:

```
> plot(treeModel)
```



Analiza wyniku:

Na podstawie zbioru Aids2 stworzono model drzewa decyzyjnego, które pozwala przewidywać, w jaki sposób dana osoba zaraziła się chorobą Aids. Zmienną która ma być predykowana jest kolumna T.categ - sposób zarażenia.

Dane zostały podzielnoe na kategorie: płeć, wiek, data śmiercim data diagnozy oraz organizacją, do której należy pacjent

```
Attribute usage:

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

Wnioski zadania 1:

Na podstawie przeprowadzonej analizy można stwierdzić najistotniesjszy element do przewidzenia sposobu zarażenia - płeć. Dla mężczyzn drugin czynnikiem jest wiek, dla kobiet natomiast data śmierci.

Zadanie 2:

Zadanie dotyczy prognozowania oceny klientów (w skali 5-punktowej, Error < 5%) urządzeń RTV AGD, określonych na Zajęciu 1. Rozwiązanie polega na użyciu pakietu mlr. Należy wybrać najlepszą metodę wśród 5 możliwych z punktu widzenia przecyzyjności. Wyniki porównywania precyzyjności metod należy przedstawić w postaci graficznej.

Wykonanie zadania:

Dołączenie pakietów (pakiet DiscriMiner musiał zostać zainstalowany lokalnie):

```
> library("mlr")
Ładowanie wymaganego pakietu: ParamHelpers
Warning message: 'mlr' is in 'maintenance-only' mode since July 2019. Future development will only happen in 'mlr3'
(<https://mlr3.mlr-org.com>). Due to the focus on 'mlr3' there might be uncaught bugs meanwhile in {mlr} - please consider
switching.
Warning messages:
1: pakiet 'mlr' został zbudowany w wersji R 4.1.3
2: pakiet 'ParamHelpers' został zbudowany w wersji R 4.1.3
> library("DiscriMiner")
Error in library("DiscriMiner") : nie ma pakietu o nazwie 'DiscriMiner'
> install.packages(DiscriMiner)
Error in install.packages : object 'DiscriMiner' not found
> install.packages("DiscriMiner")
Instalowanie pakietu w 'C:/Users/MaksioloLaptop/Documents/R/win-library/4.1'
(ponieważ 'lib' nie jest określony)
Warning in install.packages :
  package 'DiscriMiner' is not available for this version of R

A version of this package for your version of R might be available elsewhere,
see the ideas at
https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages
> install.packages("C:/Users/MaksioloLaptop/Desktop/mgr/APU/lab5/5_2/DiscriMiner_0.1-29.tar.gz", repos=NULL, type='source')
```

```
> install.packages("rFerns")
Instalowanie pakietu w 'C:/Users/MaksioloLaptop/Documents/R/win-library/4.1'
(ponieważ 'lib' nie jest określony)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/rFerns_5.0.0.zip'
Content type 'application/zip' length 211475 bytes (206 KB)
downloaded 206 KB

package 'rFerns' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\MaksioloLaptop\AppData\Local\Temp\Rtmpkf5vgb\downloaded_packages
> library("rFerns")
Warning message:
pakiet 'rFerns' został zbudowany w wersji R 4.1.3
>
```

Napisanie własnej funkcji normalizującej:

```
> normalize <- function(x)
+ {
+     return((x-min(x))/(max(x)-min(x)))
+ }
```

Zaczytanie danych lodówek:

```
> data <- read.csv("C:/Users/[redacted]/Desktop/mgr/APU/lab5/5_2/lodowki.csv")
```

Wybranie parametrów pod normalizację:

```
> data$ocena_klientow<-factor(data$ocena_klientow)
> data$poj_chlodziarki<-normalize(data$poj_chlodziarki)
> data$poj_zamrazarki<-normalize(data$poj_zamrazarki)
```

Na poniższym kroku napotkano błąd "niewspierany typ w kolumnie"

```
> zadanie=makeClassifTask(id='lodowki',
+                         data,
+                         "ocena_klientow",
+                         weights = NULL,
+                         blocking = NULL,
+                         coordinates = NULL,
+                         positive=NA_character_,
+                         fixup.data = "warn",
+                         check.data = TRUE)
Error in (function (cn, x)  :
  Unsupported feature type (character) in column 'nazwa'.
```

Nie udało się kontynuować zadania, pomimo prób naprawienia błędu.

Zapisanie widoku:

```
> save.image("C:/Users/[redacted]/Desktop/mgr/APU/lab5/5_2/Lab5_2.RData")
```

Wnioski zadania 2:

Pomimo nieudanej próby rozwiązania zadania, na zajęciach przybliżono działanie prognozowania według określonych parametrów. W przypadku zadania 2 miała to być ocena klientów. Ostatecznie zadanie miało zilustrować 5 metod możliwych I umiejscowić je odpowiednio na diagramie, w zależności od precyzyjności. Przykładowe rozwiązanie w formie

graficznej:

Wizualizacja w formie graficznej