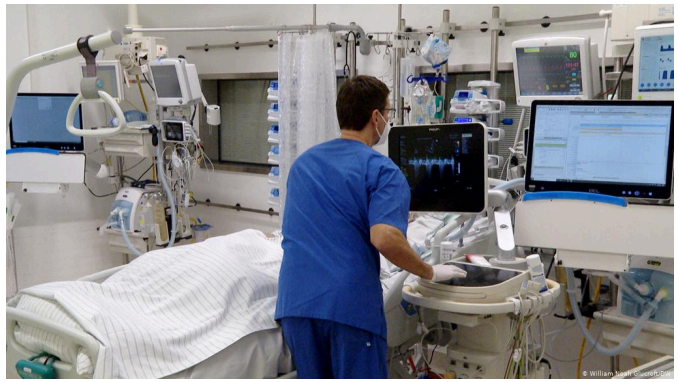
Machine Learning - 1100-ML0ENG (Ćwiczenia informatyczne Z-23/24)

Association rules II

Association rules allow us to explore the relationship between items and sets of items.

We will examine more complex relationships using the ICU dataset: a dataset about the outcomes of hospitalizations in the ICU.



```
install.packages("vcdExtra")
library(vcdExtra)
data(ICU)
summary(ICU)
```

This dataset has 200 observations and 22 attributes.

- died - refers to whether the patient died or not
- age - in years
- sex - Female, Male
- race - black, white, or other
- service - the type of ICU the patient has been admitted into (medical or surgical)
- cancer, renal, infect, and fracture - refer to the conditions the patient suffered during their stay in the ICU
  - Cancer - Wikipedia
  - Kidney - Wikipedia
  - Infection - Wikipedia
  - Bone fracture - Wikipedia
- cpr - refers to whether or not the patient underwent cardiopulmonary resuscitation

- systolic and heartrate - refer to measures of cardiac activity

- previcu - refer to whether the patient has been in the ICU previously
- admit - refer to whether the admission was elective or an emergency
- po2, ph, pco, bic, and creatin - refer to blood measures
- coma - refer to whether the patient has been in a coma
- uncons - refer to whether the patient has been unconscious at any moment during the stay in the ICU
- white - refer to race white, non-white

The race and white attributes are a bit redundant. We will therefore remove the race attribute (the fourth column).

```
ICU2 = ICU[-4]
summary(ICU2)
```

In our set we only need to have categorical variables.
Therefore, we will recode the numeric attributes into categorical attributes.

```
ICU2$age = cut(ICU2$age, breaks = 4)
ICU2$systolic = cut(ICU2$systolic, breaks = 4)
ICU2$hrtrate = cut(ICU2$hrtrate, breaks = 4)
```

Analyzing data with apriori

```
library(arules)
```

First, we need to convert the ICU to transaction format (sparse matrix named ICU_tr), before we can use it.

```
ICU_tr = as(ICU2, "transactions")
summary(ICU_tr)
```

The summary shows information, such as most frequent items, itemset distribution, and example extended item information within the dataset.

We can look inside this matrix

```
ICU_tr@data #items in "baskets"
ICU_tr@itemInfo #transactions
```

We can also use itemFrequencyPlot to visualize the most frequent items, (for instance with support over 0.5).

```
itemFrequencyPlot(ICU_tr)
itemFrequencyPlot(ICU_tr, topN=20)
itemFrequencyPlot(ICU_tr,horiz=T, support=0.5)
```

Next, we apply the Apriori algorithm to search for rules with support over 0.85 and confidence over 0.95.

```
rules <- apriori (ICU_tr,parameter = list(support = 0.85, confidence = 0.95))
inspect(rules)
```

```
rules <- apriori (ICU_tr,parameter = list(support = 0.85, confidence = 0.95,
minlen=2))
options(digits=2)
inspect(rules[1:10])
```

```
> inspect(rules[1:10])
     lhs             rhs            support confidence coverage lift count
[1]  {cancer=No}  => {creatin=<=2} 0.85    0.95       0.90     1.0  171
[2]  {pco=<=45}   => {po2=>60}     0.86    0.96       0.90     1.0  172
[3]  {pco=<=45}   => {ph=>=7.25}   0.88    0.97       0.90     1.0  175
[4]  {renal=No}   => {cpr=No}      0.86    0.95       0.90     1.0  172
[5]  {renal=No}   => {ph=>=7.25}   0.86    0.95       0.90     1.0  172
[6]  {renal=No}   => {creatin=<=2} 0.88    0.97       0.90     1.0  176
[7]  {po2=>60}    => {ph=>=7.25}   0.88    0.96       0.92     1.0  176
[8]  {po2=>60}    => {creatin=<=2} 0.88    0.95       0.92     1.0  175
[9]  {uncons=No}  => {coma=None}   0.92    1.00       0.92     1.1  185
[10] {coma=None}  => {uncons=No}   0.92    1.00       0.92     1.1  185
```

With high confidence and support,

- the absence of a cancer is associated with creatin levels lower than or equal to 2.
- Low arterial concentration of carbonic oxide (<= 45) is associated with high blood concentration of oxygen (>= 60).
- Patients who did not have a history of renal failure did not need CPR, had a blood ph higher or equal to 7.25, and had a creatin level lower or equal to 2.
- Patients with a blood concentration of oxygen higher than or equal to 60 had a blood pH higher than or equal to 7.25 and a creatin level equal to or lower than 2.

Next, let's looking at what the antecedents of death in the ICU.

For this analysis, we will set the rhs parameter of the appearance argument to died=Yes. We will use lower confidence and support thresholds for this
analysis as follows:

```
rulesDeath <- apriori(ICU_tr,
  parameter = list(support=0.1, confidence=0.3),
  appearance = list(rhs = c("died=Yes"), default="lhs"))
inspect(rulesDeath[1:10])
```

```
> inspect(rulesDeath[1:10])
     lhs                                                  rhs          support con
[1]  {infect=Yes,admit=Emergency}                      => {died=Yes} 0.12    0.3
[2]  {service=Medical,white=Non-white}                 => {died=Yes} 0.12    0.3
[3]  {infect=Yes,admit=Emergency,white=Non-white}      => {died=Yes} 0.12    0.3
[4]  {cancer=No,infect=Yes,admit=Emergency}            => {died=Yes} 0.12    0.3
[5]  {infect=Yes,admit=Emergency,pco=<=45}             => {died=Yes} 0.10    0.3
[6]  {infect=Yes,admit=Emergency,po2=>60}              => {died=Yes} 0.10    0.3
[7]  {infect=Yes,admit=Emergency,fracture=No}          => {died=Yes} 0.11    0.3
[8]  {infect=Yes,admit=Emergency,ph=>=7.25}            => {died=Yes} 0.10    0.3
[9]  {cancer=No,infect=Yes,white=Non-white}            => {died=Yes} 0.11    0.3
[10] {infect=Yes,po2=>60,white=Non-white}              => {died=Yes} 0.10    0.3
```

We can see that 34 percent of patients who were admitted in emergency and had an infection died in the ICU, as did 30 percent of patients who were non-white and whose ICU admission service was medical.

Last modified: piątek, 18 listopada 2022, 6:58

Uniwersytetu Łódzkiego obowiązuje nazwa użytkownika i hasło wykorzystywane podczas logowania się do systemu USOSweb.