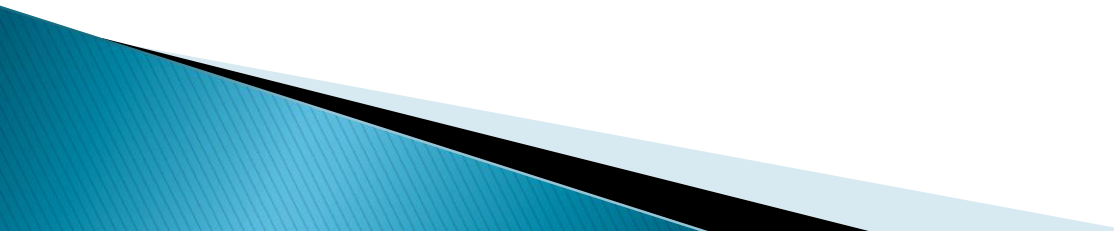


PROJEKT PREZENTACJA 4

- ▶ Wiktoria Śmigecka 165614
- ▶ Aleksandra Żmijewska 165232

Rozwój aplikacji internetowych w medycynie
Hurtownie i eksploracja danych

Plan prezentacji

1. Temat i cel projektu.
 2. Zgromadzenie danych.
 3. Przetwarzanie i oczyszczanie danych.
 4. Algorytm Apriori.
 5. Naiwny Klasyfikator Bayesa.
 6. Drzewa decyzyjne.
- 

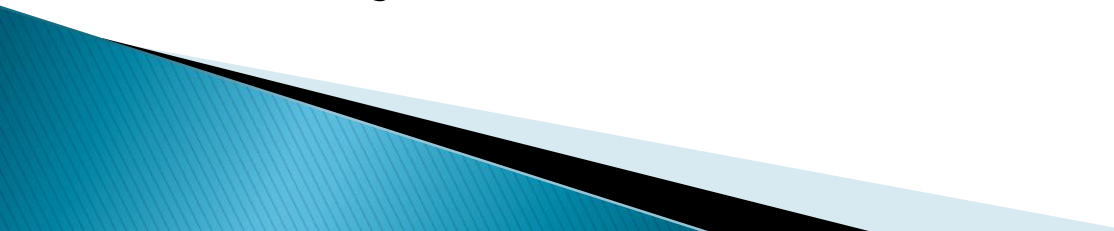
Temat projektu:

Wnioskowanie na temat ryzyka choroby

Cel projektu:

Celem projektu jest gromadzenie danych oraz analiza czynników choroby Downa.

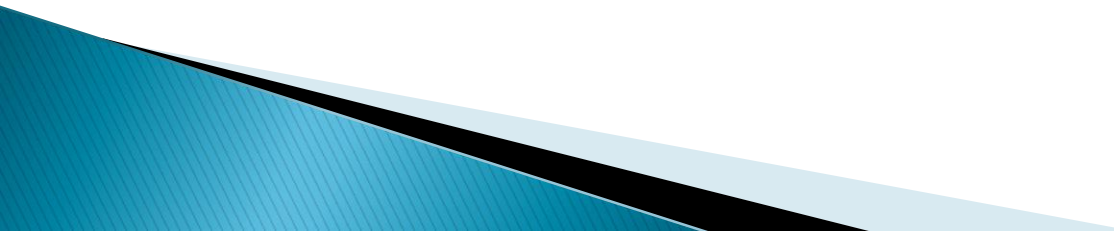
Celem projektu jest także przedstawienie generatora informacji o chorobie Downa i wizualizacja informacji



Efekty projektu:

- ▶ Efektem projektu powinny być wygenerowane i opisane reguły wiedzy wraz z ich odpowiednią oceną
- ▶ Efektem projektu planowane jest oprogramowanie diagnostyczne składające się z:
 - > Modułu prezentacji graficznej danych

Wykonanie projektu

- ▶ Wykonany na serwerze AWS, wykorzystano program Putty i system Ubuntu
 - ▶ Interfejsy wykonane za pomocą HTML'a
 - ▶ Wykorzystano moduł Python 2.7
 - ▶ Baza danych SQLite3
- 

Baza danych

```
sqlite> SELECT id, wiek_matki, wiek_ojca, ilosc_dzieci_w_rodzinie, wystepowanie, kontakt_z_prom FROM osoba  
...> ;
```

id	wiek_matki	wiek_ojca	ilosc_dzieci_w_rodzinie	wystepowanie	kontakt_z_prom
1	45	56	0	Nie	Nie
2	50	70	7	Tak	Tak
3	24	29	2	Tak	Tak
4	30	31	1	Tak	Nie
5	27	30	4	Tak	Tak

```
sqlite> █
```

```
sqlite> SELECT id, kontakt_z_sub, leki, ciaza, odstep, krotka_szyja FROM osoba  
...> ;
```

id	kontakt_z_sub	leki	ciaza	odstep	krotka_szyja
1	Nie	Nie	nie dotyczy	nie dotyczy	1
2	Tak	Tak	Nie	Cztery mies	0
3	Nie	Nie	Nie	nie dotyczy	0
4	Nie	Tak	Tak	Pół roku	0
5	Nie	Tak	Nie	Dwa miesiąc	0

```
sqlite> █
```

Zgromadzenie danych

Kwestionariusz został udostępniony w gronie rodziców dzieci z chorobą Downa a także na portalach społecznościowych.

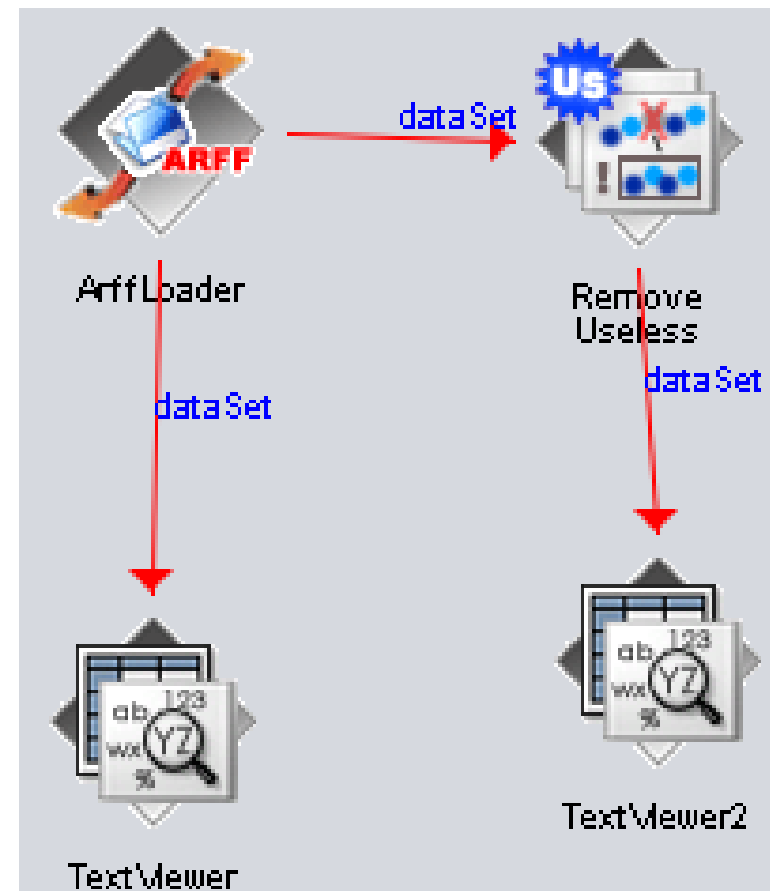
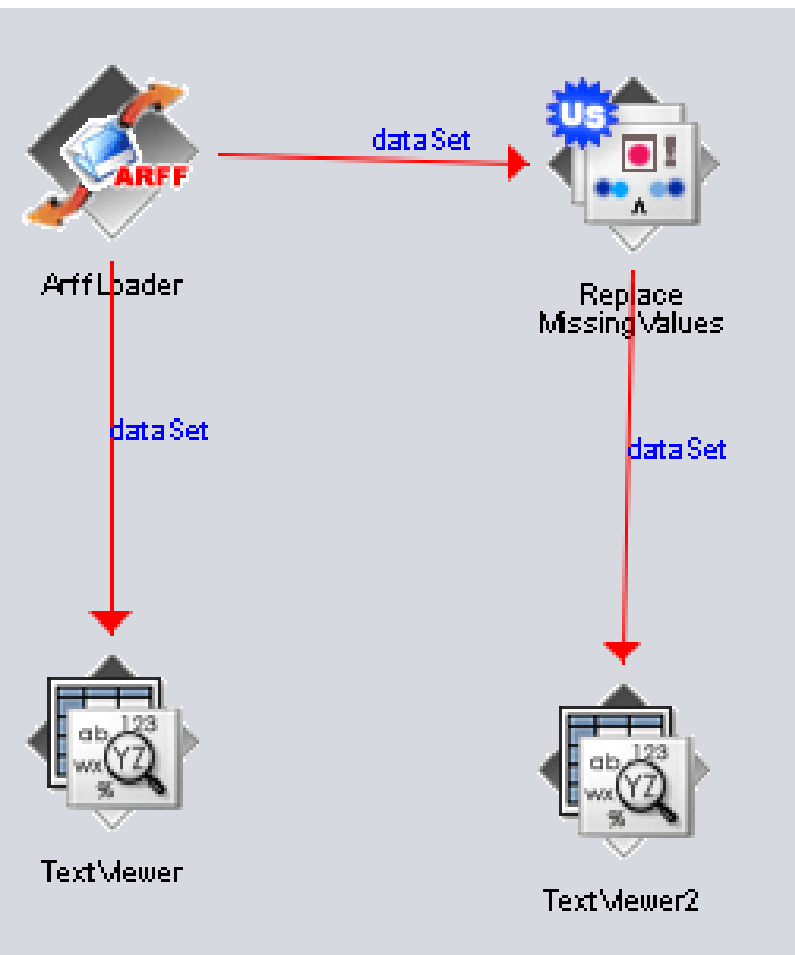
Został wypełniony przez 42 osoby. Po oczyszczeniu danych i usunięciu niepełnych danych otrzymano 40 rekordów.

Do analizy danych wykorzystano oprogramowanie WEKA.



Wstępne przetwarzanie danych oczyszczanie danych

Usunięto niepełne rekordy a także bezużyteczne dane



Fragment pliku ARFF

```
@relation osoba-2.1
```

```
@attribute ilosc_dzieci_w_rodzinie {ilosc_2,ilosc_0,ilosc_1,ilosc_5,ilosc_3,ilosc_10}
```

```
@attribute wiek_ojca {wiek_ojca_18,wiek_ojca_27,wiek_ojca_50,wiek_ojca_7,wiek_ojca_38,wiek_ojca_37,wiek_ojca_29,wiek_ojca_35,wiek_ojca_24,wiek_ojca_45,wwiek_ojca_10}
```

```
@attribute wiek_matki {wiek_matki_18,wiek_matki_19,wiek_matki_50,wiek_matki_44,wiek_matki_40,wiek_matki_28,wiek_matki_35,wiek_matki_25,wiek_matki_33,wiek_matki_24}
```

```
@attribute wystepowanie {wystepowanie_Tak,wystepowanie_Nie}
```

```
@attribute kontakt_z_prom {kontakt_z_prom_Nie,kontakt_z_prom_Tak}
```

```
@attribute kontakt_z_sub {kontakt_z_sub_Nie,kontakt_z_sub_Tak}
```

```
@attribute leki {leki_Nie,leki_Tak}
```

```
@attribute ciaza {ciaza_nie_dotyczy,ciaza_Nie,ciaza_Tak}
```

```
@attribute odstep_leki {odstep_nie_dotyczy,'odstep_Ponad rok','odstep_Trzy miesiace',odstep_Pol_roku,odstep_Trzy_miesiace,'odstep_nie dotyczy',odstep_Ponad_rok}
```

```
@attribute krotka_szyja {krotka_szyja_1,krotka_szyja_0}
```

```
@attribute male_uszy {uszy_0,uszy_1}
```

```
@attribute plaska_twarz {plaska_twarz_0,plaska_twarz_1}
```

```
@attribute krotkie_dlonie {krotkie_dlonie0,krotkie_dlonie1}
```

```
@attribute skosno_ustawione_powieki {powieki_0,powieki_1}
```

```
@attribute slaby_wzrok {slaby_wzrok_0,slaby_wzrok_1}
```

```
@attribute opozniony_rozwoj {opozniony_rozwoj_1,opozniony_rozwoj_0}
```

```
@attribute wady_serca {serce_0,serce_1}
```

```
@attribute wady_sluchu {sluch_0,sluch_1}
```

```
@attribute wady_wzroku {wzrok_0,wzrok_1}
```

```
@attribute obnizona_odpornosc {obnizona_odpornosc_0,obnizona_odpornosc_1}
```

```
@data
```

```
ilosc_2,wiek_ojca_18,wiek_matki_18,wystepowanie_Tak,kontakt_z_prom_Nie,kontakt_z_sub_Nie,leki_Nie,ciaza_nie_dotyczy,odstep_nie_dotyczy,krotka_szyja_1,uszy_0,p
```

```
ilosc_0,wiek_ojca_27,wiek_matki_19,wystepowanie_Nie,kontakt_z_prom_Nie,kontakt_z_sub_Nie,leki_Nie,ciaza_nie_dotyczy,odstep_nie_dotyczy,krotka_szyja_0,uszy_1,p
```

```
ilosc_2,wiek_ojca_18,wiek_matki_18,wystepowanie_Tak,kontakt_z_prom_Nie,kontakt_z_sub_Nie,leki_Nie,ciaza_nie_dotyczy,odstep_nie_dotyczy,krotka_szyja_1,uszy_0,p
```

```
ilosc_1,wiek_ojca_50,wiek_matki_50,wystepowanie_Nie,kontakt_z_prom_Nie,kontakt_z_sub_Nie,leki_Nie,ciaza_nie_dotyczy,odstep_nie_dotyczy,krotka_szyja_1,uszy_0,p
```

```
ilosc_5,wiek_ojca_50,wiek_matki_44,wystepowanie_Nie,kontakt_z_prom_Tak,kontakt_z_sub_Tak,leki_Tak,ciaza_nie_dotyczy,odstep_nie_dotyczy,krotka_szyja_0,uszy_0,p
```

```
ilosc_2,wiek_ojca_7,wiek_matki_40,wystepowanie_Tak,kontakt_z_prom_Tak,kontakt_z_sub_Tak,leki_Tak,ciaza_Nie,'odstep_Ponad rok',krotka_szyja_1,uszy_0,plaska_twa
```

```
ilosc_1,wiek_ojca_38,wiek_matki_28,wystepowanie_Nie,kontakt_z_prom_Tak,kontakt_z_sub_Nie,leki_Nie,ciaza_nie_dotyczy,odstep_nie_dotyczy,krotka_szyja_1,uszy_0,p
```

```
ilosc_2,wiek_ojca_37,wiek_matki_35,wystepowanie_Nie,kontakt_z_prom_Nie,kontakt_z_sub_Nie,leki_Nie,ciaza_nie_dotyczy,odstep_nie_dotyczy,krotka_szyja_1,uszy_0,p
```

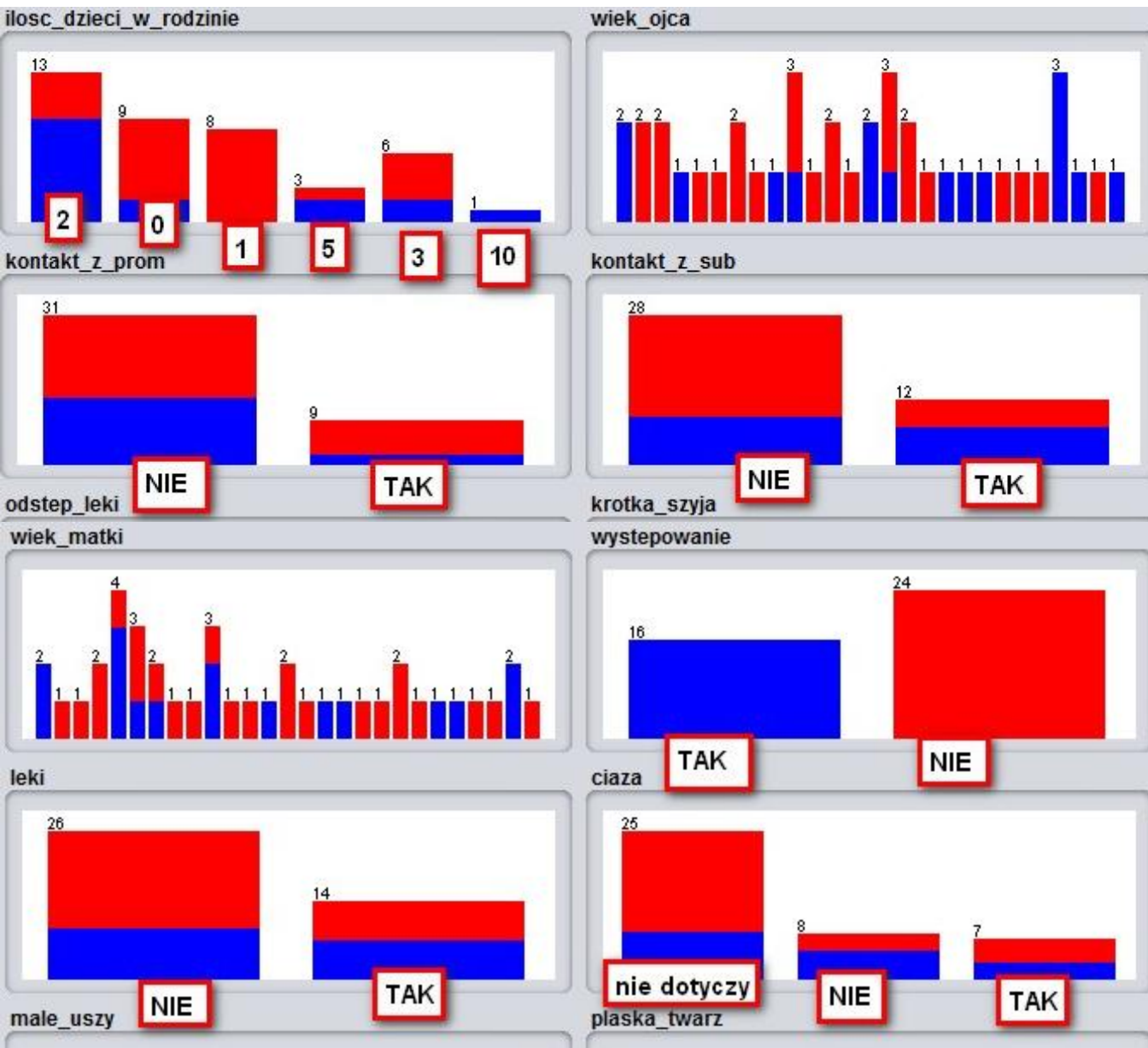
```
ilosc_2,wiek_ojca_29,wiek_matki_25,wystepowanie_Nie,kontakt_z_prom_Nie,kontakt_z_sub_Nie,leki_Tak,ciaza_Nie,'odstep_Trzy miesiace',krotka_szyja_0,uszy_0,plask
```

```
ilosc_1,wiek_ojca_35,wiek_matki_33,wystepowanie_Nie,kontakt_z_prom_Nie,kontakt_z_sub_Nie,leki_Nie,ciaza_nie_dotyczy,odstep_nie_dotyczy,krotka_szyja_0,uszy_0,p
```

```
ilosc_2,wiek_ojca_24,wiek_matki_24,wystepowanie_Tak,kontakt_z_prom_Nie,kontakt_z_sub_Nie,leki_Nie,ciaza_Nie,odstep_nie_dotyczy,krotka_szyja_1,uszy_0,plaska_tw
```

```
ilosc_3,wiek_ojca_45,wiek_matki_44,wystepowanie_Nie,kontakt_z_prom_Nie,kontakt_z_sub_Nie,leki_Nie,ciaza_Nie,odstep_nie_dotyczy,krotka_szyja_1,uszy_1,plaska tw
```

Diagramy danych



występowanie
choroby w rodzinie

niewystępowanie
choroby w rodzinie

Algorytm apriori

Minimum support: 0.6 (24 instances)

Minimum metric <confidence>: 0.5

Number of cycles performed: 8

1. leki=leki_Nie 26 ==> odstep_leki=odstep_nie_dotyczy 26

<conf:(1)> lift:(1.21) lev:(0.11) [4] conv:(4.55)

2. ciaza=ciaza_nie_dotyczy 25 ==> odstep_leki=odstep_nie_dotyczy 25

<conf:(1)> lift:(1.21) lev:(0.11) [4] conv:(4.38)

3. kontakt_z_sub=kontakt_z_sub_Nie leki=leki_Nie 24 ==> odstep_leki=odstep_nie_dotyczy 24

<conf:(1)> lift:(1.21) lev:(0.11) [4] conv:(4.2)

4. leki=leki_Nie ciaza=ciaza_nie_dotyczy 24 ==> odstep_leki=odstep_nie_dotyczy 24

<conf:(1)> lift:(1.21) lev:(0.11) [4] conv:(4.2)

5. kontakt_z_sub=kontakt_z_sub_Nie 28 ==> odstep_leki=odstep_nie_dotyczy 27

<conf:(0.96)> lift:(1.17) lev:(0.1) [3] conv:(2.45)

6. ciaza=ciaza_nie_dotyczy 25 ==> leki=leki_Nie 24

<conf:(0.96)> lift:(1.48) lev:(0.19) [7] conv:(4.38)

7. kontakt_z_prom=kontakt_z_prom_Nie kontakt_z_sub=kontakt_z_sub_Nie 25 ==>
odstep_leki=odstep_nie_dotyczy 24

<conf:(0.96)> lift:(1.16) lev:(0.08) [3] conv:(2.19)

8. ciaza=ciaza_nie_dotyczy odstep_leki=odstep_nie_dotyczy 25 ==> leki=leki_Nie 24

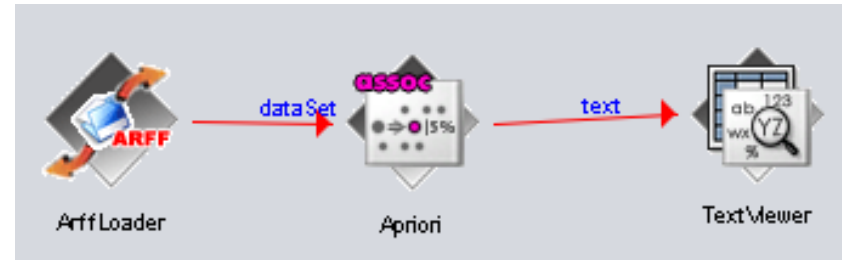
<conf:(0.96)> lift:(1.48) lev:(0.19) [7] conv:(4.38)

9. ciaza=ciaza_nie_dotyczy 25 ==> leki=leki_Nie odstep_leki=odstep_nie_dotyczy 24

<conf:(0.96)> lift:(1.48) lev:(0.19) [7] conv:(4.38)

10. leki=leki_Nie 26 ==> kontakt_z_sub=kontakt_z_sub_Nie 24

<conf:(0.92)> lift:(1.32) lev:(0.15) [5] conv:(2.6)

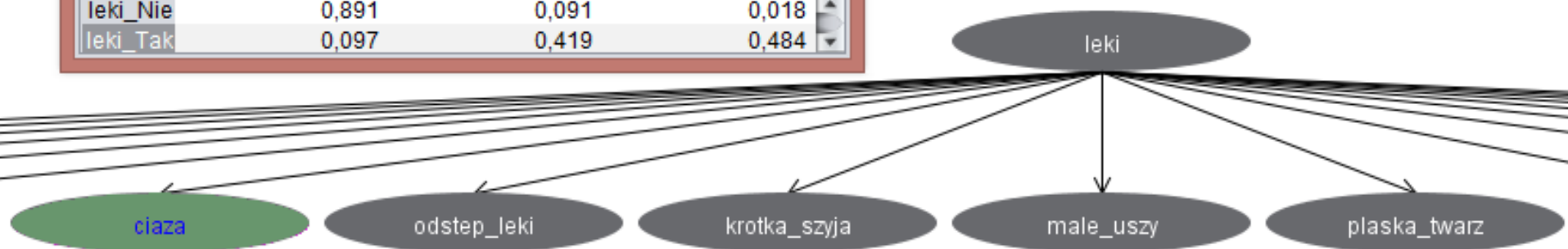


Graf klasyfikatora Bayesa



Probability Distribution Table For ciężka

leki	ciężka_nie_dotyczy	ciężka_Nie	ciężka_Tak
leki_Nie	0,891	0,091	0,018
leki_Tak	0,097	0,419	0,484



Klasifikator Bayesa

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	33	82.5	%
Incorrectly Classified Instances	7	17.5	%
Kappa statistic	0.6089		
Mean absolute error	0.1374		
Root mean squared error	0.3026		
Relative absolute error	29.8411	%	
Root relative squared error	62.9654	%	
Total Number of Instances	40		

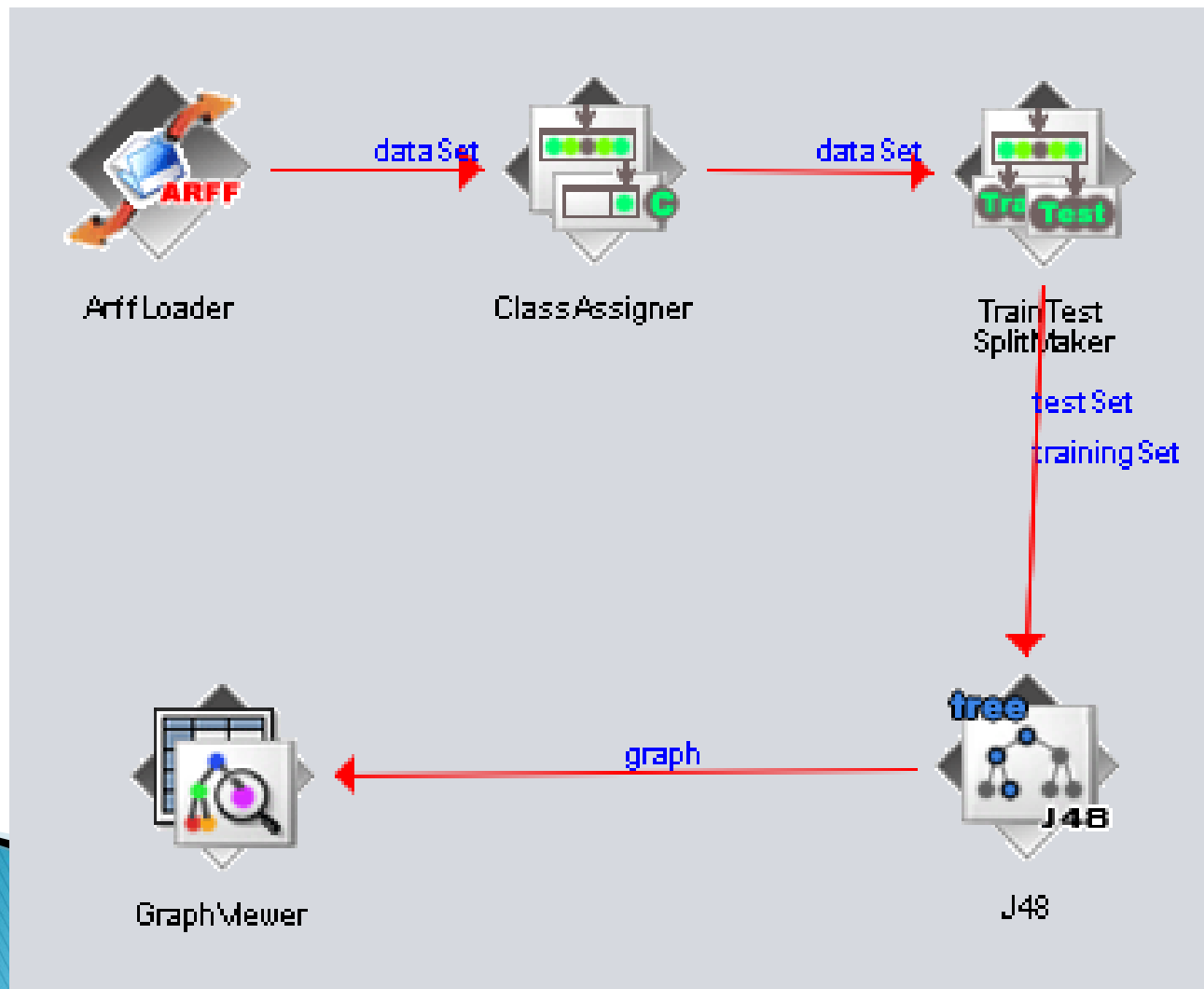
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,885	0,286	0,852	0,885	0,868	0,610	0,953	0,975	leki_Nie
	0,714	0,115	0,769	0,714	0,741	0,610	0,953	0,931	leki_Tak
Weighted Avg.	0,825	0,226	0,823	0,825	0,823	0,610	0,953	0,960	

=== Confusion Matrix ===

```
a  b  <-- classified as
23  3 |  a = leki_Nie
 4 10 |  b = leki_Tak
```

Sposób generowania drzew decyzyjnych



Opis drzewa decyzyjnego

J48 pruned tree

leki = leki_Nie: kontakt_z_sub_Nie (26.0/2.0)

leki = leki_Tak

| krotka_szyja = krotka_szyja_1: kontakt_z_sub_Tak (7.0)

| krotka_szyja = krotka_szyja_0

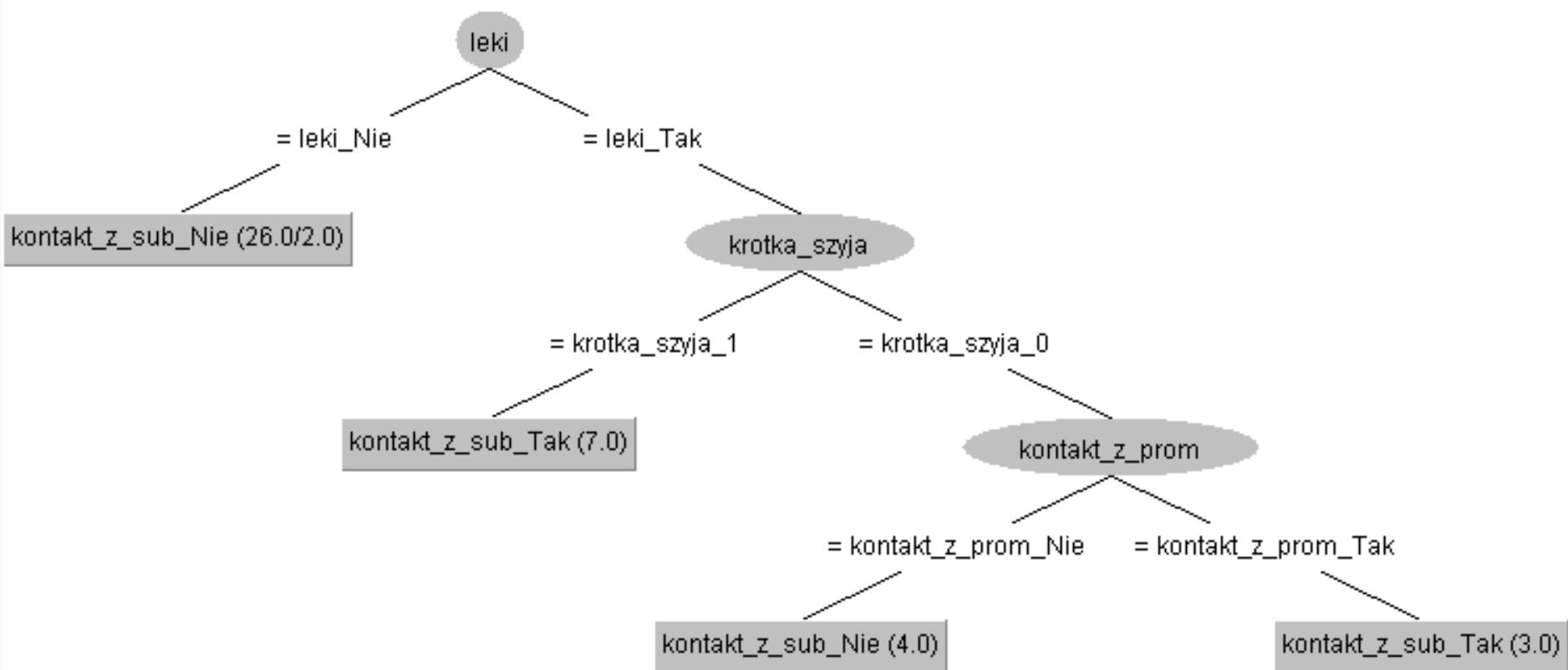
| | kontakt_z_prom = kontakt_z_prom_Nie: kontakt_z_sub_Nie (4.0)

| | kontakt_z_prom = kontakt_z_prom_Tak: kontakt_z_sub_Tak (3.0)

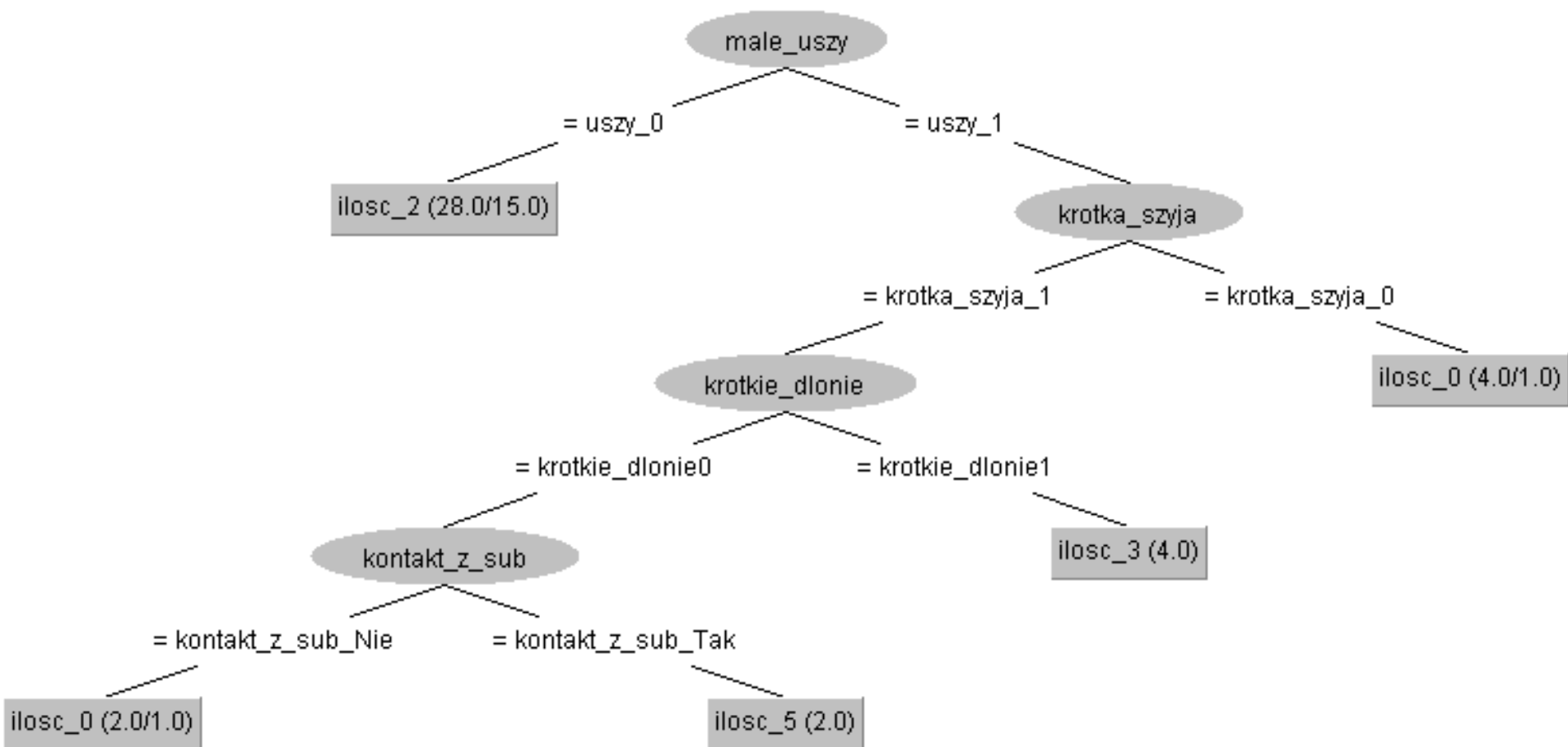
Number of Leaves : 4

Size of the tree : 7

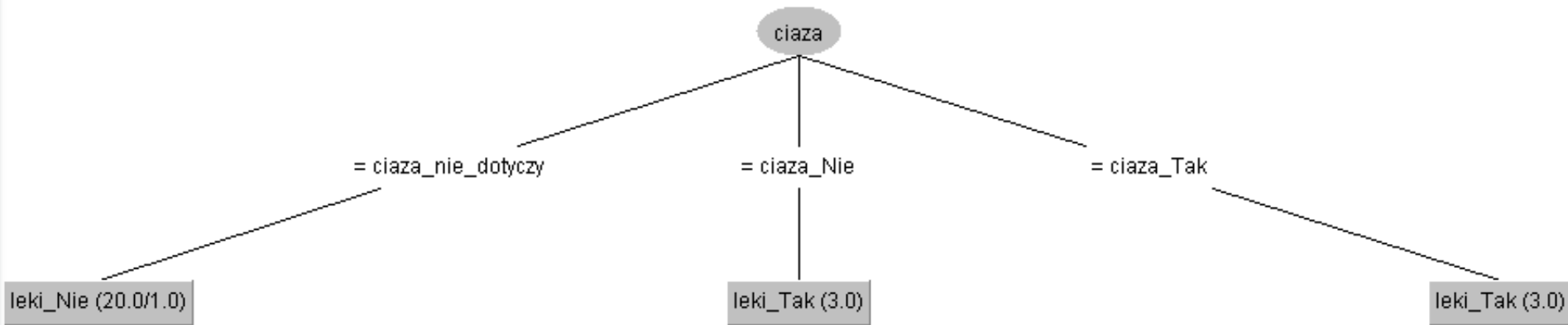
Drzewo decyzyjne



Drzewo decyzyjne



Drzewo decyzyjne



Dziękujemy za uwagę