

1. Zbiór *dlbcl*

Selekcja atrybutów na zbiorze treningowym *dlbcl_training.csv*

Metoda **ReliefAttributeEval** -M -1 -D 1 -K 10

Zostały wybrane następujące cechy (pattern):

Z11793_at|U50527_s_at|X02152_at|V00594_s_at|M91196_at|L19686_rna1_at|S73591_at|U72935_cds3_s_at|HG1980-HT2023_at|L17131_rna1_at|class

Ranked attributes:

| | | |
|-------|------|------------------|
| 0.272 | 2550 | Z11793_at |
| 0.248 | 1727 | U50527_s_at |
| 0.243 | 2000 | X02152_at |
| 0.243 | 1988 | V00594_s_at |
| 0.212 | 1279 | M91196_at |
| 0.209 | 768 | L19686_rna1_at |
| 0.204 | 1345 | S73591_at |
| 0.203 | 1860 | U72935_cds3_s_at |
| 0.196 | 450 | HG1980-HT2023_at |
| 0.195 | 759 | L17131_rna1_at |

Selekcja atrybutów metodą **OneRAttributeEval** -S 1 -F 10 -B 6 :

Pattern:

M12963_s_at|U25826_at|J02645_at|HG2874-HT3018_at|HG1980-HT2023_at|X62078_at|L36720_at|X03934_at|X65550_at|D38751_at|class

Ranked attributes:

| | | |
|--------|------|------------------|
| 91.803 | 907 | M12963_s_at |
| 88.525 | 1562 | U25826_at |
| 88.525 | 584 | J02645_at |
| 86.885 | 483 | HG2874-HT3018_at |

86.885 450 HG1980-HT2023_at
86.885 2208 X62078_at
86.885 839 L36720_at
86.885 2016 X03934_at
86.885 2238 X65550_at
86.885 232 D38751_at

Selected attributes: 907,1562,584,483,450,2208,839,2016,2238,232 : 10

Selekcja atrybutów metodą [InfoGainAttributeEval](#) :

Pattern:

HG1980-HT2023_at|X02152_at|J04031_at|HG4074-
HT4344_at|L36720_at|V00594_at|U63743_at|L08187_at|X62078_at|X74795_at|class

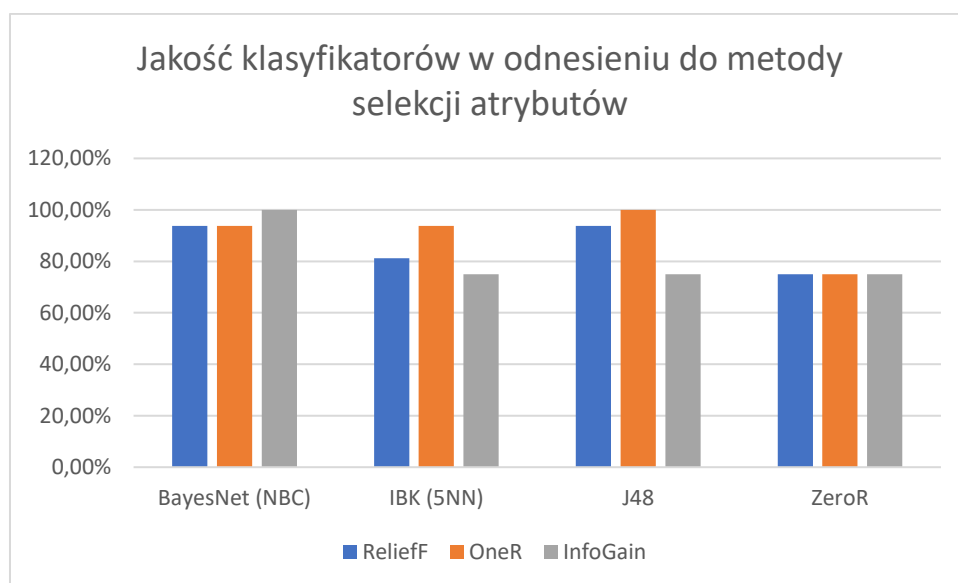
Ranked attributes:

0.544 450 HG1980-HT2023_at
0.453 2000 X02152_at
0.453 617 J04031_at
0.434 534 HG4074-HT4344_at
0.434 839 L36720_at
0.434 1987 V00594_at
0.429 1807 U63743_at
0.408 705 L08187_at
0.407 2208 X62078_at
0.407 2317 X74795_at

Selected attributes: 450,2000,617,534,839,1987,1807,705,2208,2317 : 10

Wyniki klasyfikacji dla wybranych dziesięciu atrybutów dla zbioru testowego *dlbcl_testing.csv*:

| Selekcja atrybutów | Metody klasyfikacji | | | |
|--------------------|---------------------|-----------|--------|-------|
| | BayesNet (NBC) | IBK (5NN) | J48 | ZeroR |
| ReliefF | 93,75% | 81,25% | 93,75% | 75% |
| OneR | 93,75% | 93,75% | 100% | 75% |
| InfoGain | 100% | 75% | 75% | 75% |



Wnioski: dla zbioru *dlbcl_testing.csv* filtr wyboru cech istotnych ma znaczenie, gdyż testowane klasyfikatory (oprócz ZeroR) dla różnych metod selekcji atrybutów zwracają różne wyniki.

2. Zbiór *prostate*

Selekcja atrybutów na zbiorze treningowym *prostate_training.csv*

Metoda **ReliefAttributeEval** -M -1 -D 1 -K 10

Pattern:

37639_at|41468_at|32598_at|40282_s_at|38634_at|37366_at|41706_at|32109_at|37394_at|38827_at|class

Ranked attributes:

| | | |
|-------|------|------------|
| 0.356 | 1226 | 37639_at |
| 0.258 | 1945 | 41468_at |
| 0.231 | 432 | 32598_at |
| 0.222 | 1737 | 40282_s_at |
| 0.209 | 1409 | 38634_at |
| 0.196 | 1181 | 37366_at |
| 0.182 | 1970 | 41706_at |
| 0.179 | 342 | 32109_at |
| 0.177 | 1187 | 37394_at |
| 0.173 | 1469 | 38827_at |

Selected attributes: 1226,1945,432,1737,1409,1181,1970,342,1187,1469 : 10

Selekcja atrybutów metodą **OneRAttributeEval** -S 1 -F 10 -B 6 :

Pattern:

37639_at|32598_at|41706_at|38028_at|34840_at|38827_at|38291_at|37366_at|35276_at|38059_g_at|class

Ranked attributes:

| | | |
|---------|------|----------|
| 89.0244 | 1226 | 37639_at |
| 87.8049 | 432 | 32598_at |

| | |
|---------|-----------------|
| 87.8049 | 1970 41706_at |
| 85.3659 | 1282 38028_at |
| 84.1463 | 728 34840_at |
| 84.1463 | 1469 38827_at |
| 84.1463 | 1333 38291_at |
| 84.1463 | 1181 37366_at |
| 82.9268 | 783 35276_at |
| 82.9268 | 1293 38059_g_at |

Selected attributes: 1226,432,1970,1282,728,1469,1333,1181,783,1293 : 10

Selekcja atrybutów metodą InfoGainAttributeEval :

Pattern:

32598_at|41706_at|37639_at|37366_at|1831_at|38028_at|38057_at|38291_at|38634_at|41468_at|class

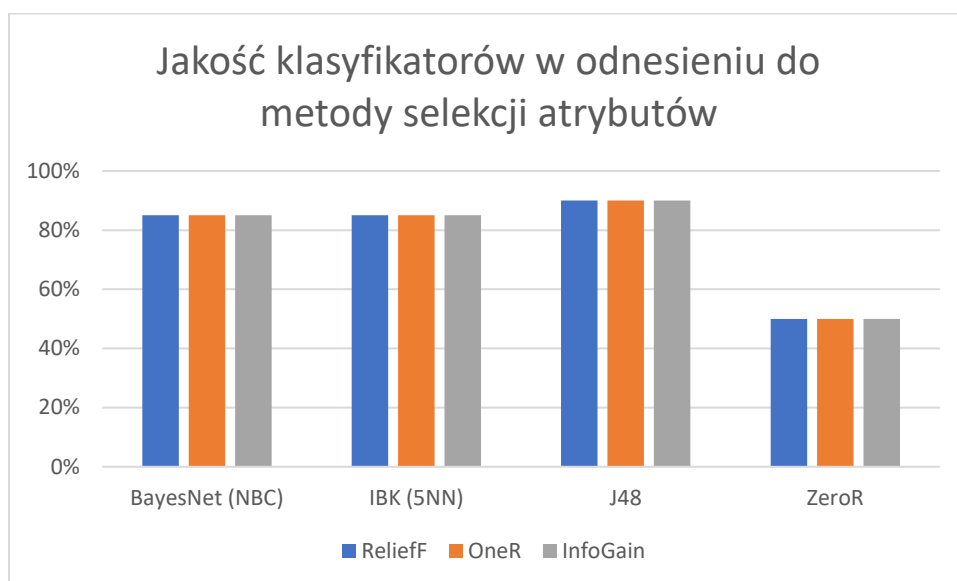
Ranked attributes:

| | |
|-------|---------------|
| 0.613 | 432 32598_at |
| 0.589 | 1970 41706_at |
| 0.542 | 1226 37639_at |
| 0.511 | 1181 37366_at |
| 0.477 | 152 1831_at |
| 0.473 | 1282 38028_at |
| 0.463 | 1292 38057_at |
| 0.446 | 1333 38291_at |
| 0.446 | 1409 38634_at |
| 0.432 | 1945 41468_at |

Selected attributes: 432,1970,1226,1181,152,1282,1292,1333,1409,1945 : 10

Wyniki klasyfikacji dla wybranych dziesięciu atrybutów dla zbioru testowego *prostate_testing.csv*:

| Selekcja atrybutów | Metody klasyfikacji | | | |
|--------------------|---------------------|-----------|-----|-------|
| | BayesNet (NBC) | IBK (5NN) | J48 | ZeroR |
| ReliefF | 85% | 85% | 90% | 50% |
| OneR | 85% | 85% | 90% | 50% |
| InfoGain | 85% | 85% | 90% | 50% |



Wnioski: dla zbioru *prostate_testing.csv* filtr wyboru cech istotnych nie ma znaczenia (dla testowanych filtrów), gdyż wszystkie klasyfikatory dla różnych metod selekcji atrybutów zwracają jednakowe wyniki.