Przykład eksploracji danych
Case 1.X

JERZY STEFANOWSKI

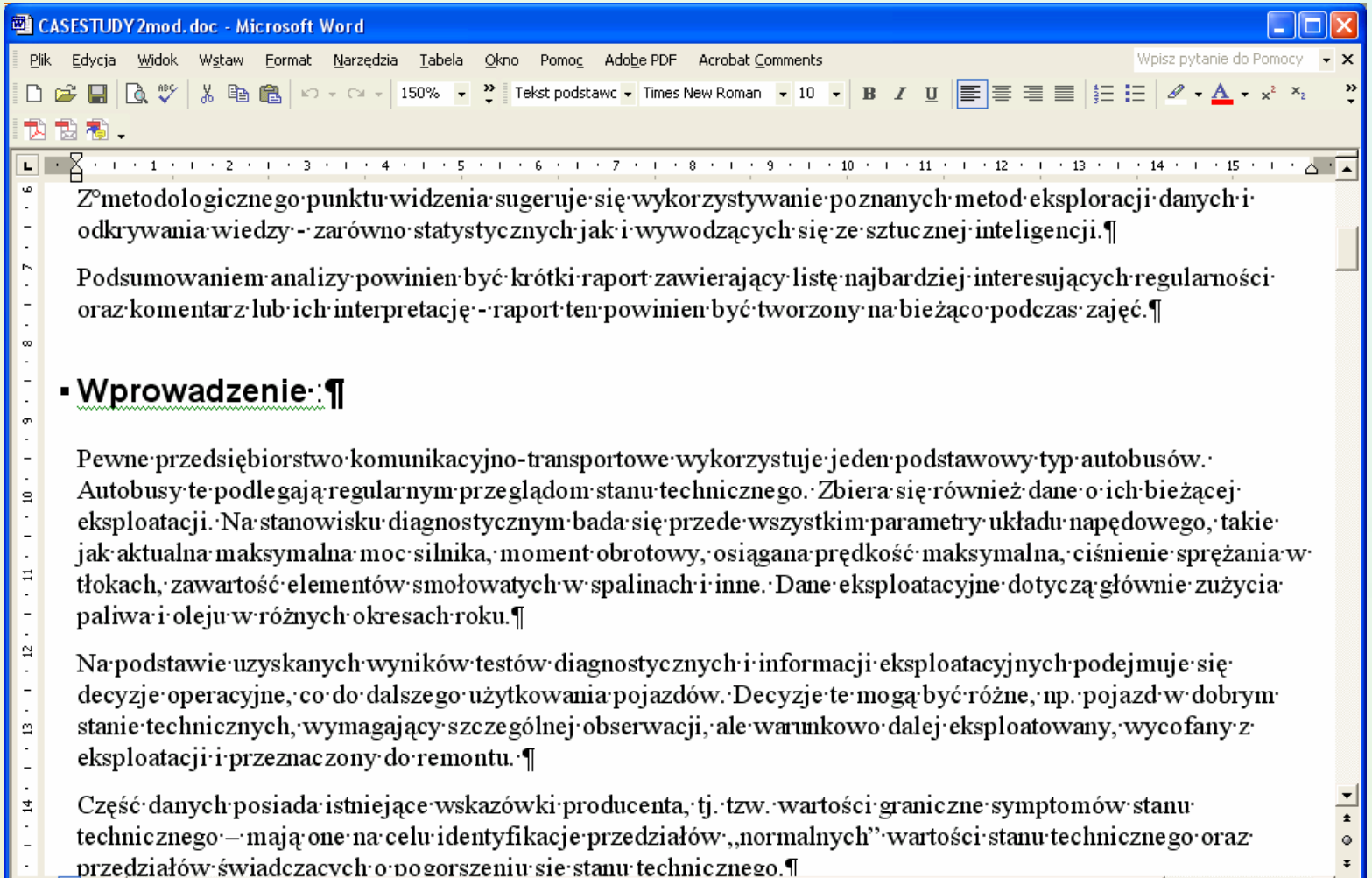
TPD – Zaawansowana eksploracja danych
edycja 2009/2010

Plan

1. Przykładowe studium przypadki
2. Analiza opisu przypadku
3. Ustalenie celu analizy i scenariusza postępowania
4. Ocena poprawności danych – „czyszczenie”
5. Badanie jakości danych (współzależność)
6. Ocena ważności atrybutów
7. Odkrywanie wiedzy klasyfikacyjnej
    1. Różne podejścia
    2. Ocena zdolności klasyfikacyjnej
    3. Możliwości interpretacji wiedzy
8. Wymagania do sprawozdania

Analiza diagnostycznej bazy danych

- Problem dotyczy analizy stanu technicznego autobusów używanych przez jedno z przedsiębiorstw w Polsce.



Analiza diagnostycznej bazy danych

- Bada się stan techniczny 80 autobusów tego samego typu (dokładnie ich silników) na podstawie symptomów stanu technicznego - parametrów pochodzących z okresowych badań diagnostycznych
  - Pierwsza klasyfikacja D1: autobusy są podzielone na dwie klasy: dobry i zły stan techniczny pojazdu
  - Możliwa jest druga klasyfikacja D2 + stan przejściowy
- Cel analizy
  - Ocenia się jakość diagnostyczną symptomów stanu technicznego
    - (pośrednio ocena przydatności tzw. wartości granicznych)
  - Ocena ważności poszczególnych symptomów
    - Ewentualność rankingu lub selekcji
  - Poszukuje się zależności pomiędzy wartościami najistotniejszych w tych symptomów a przydziałem do klas
  - Konstruuje się klasyfikator stanu technicznego

Oryginalny format danych (isf)

autobusy.isf - Notatnik

Plik Edycja Format Widok Pomoc

```
**ATTRIBUTES
A0:      (omit) | number of example
MaxSpeed:      (continuous) | km/h
Compr_preasure:      (continuous) | Mpa
blacking:      (continuous) | blacking components in exhaust gas [%]
torque:      (continuous) | Nm
summer_cons:      (continuous) | summer fuel consumption l/100km
winter_cons:      (continuous) | winter fuel consumption l/100km
oil_cons:      (continuous) | oil consumption l/1000km
horsepower:      (continuous) | max horesepowewr Km
D1:      [1,2] | technical condtion of a vehicle 1 - good, 2 - bad
D2:      (omit) | [1,2,3]
decision: D1

**EXAMPLES
1,   90,   2.52,   38,   481,   21.8,   26.4,   0.7,   145,    1,   1
2,   76,   2.11,   70,   420,   22.0,   25.5,   2.7,   110,    2,   3
3,   63,   1.96,   82,   400,   22.0,   24.8,   3.7,   101,    2,   3
4,   90,   2.48,   49,   477,   21.9,   25.1,   1.0,   138,    1,   1
5,   85,   2.45,   52,   460,   21.8,   25.2,   1.4,   130,    1,   2
6,   72,   2.20,   73,   425,   23.1,   27.4,   2.8,   112,    2,   3
7,   88,   2.50,   50,   480,   21.6,   24.7,   1.1,   140,    1,   1
8,   87,   2.48,   56,   465,   22.8,   27.6,   1.4,   135,    1,   1
9,   90,   2.56,   16,   486,   26.5,   27.3,   0.2,   150,    1,   1
10,  60,   1.95,   95,   400,   23.3,   24.8,   4.4    96 ,    2,   3
11,  80,   2.41,   60,   451,   21.7,   26.1,   1.7,   125,    1,   2
12,  78,   2.4 ,   63,   448,   21.8,   26.0,   1.9,   120,    1,   2
13,  90,   2.58,   26,   482,   22.4,   24.5,   0.4,   148,    1,   1
14,  62,   1.98,   93,   400,   22   ,   28.4,   3.9,   100,    2,   3
15,  82,   2.5 ,   54,   461,   22.0,   26.3,   1.4,   132,    1,   2
```

Analiza nagłówka

- 80 obserwacji (autobusów)
- 8 atrybutów
  - max speed km/h
  - Compresion preasure Mpa
  - blacking components in exhaust gas [%]
  - torque Nm
  - summer fuel consumption l/100km
  - winter fuel consumption l/100km
  - oil consumption l/1000km
  - max horesepower kM
- D1: [1,2] | technical conditions of a vehicle: 1 - good, 2 - bad
- D2: [1,2,3]

Wartości graniczne symptomów

CASESTUDY2mod.doc - Microsoft Word

Plik Edycja Widok Wstaw Format Narzędzia Tabela Okno Pomoc Adobe PDF Acrobat Comments

Wpisz pytanie do Pomocy

150% Tekst podstawc Times New Roman 10

- Symptomy· mające· w· oryginale· wartości· liczbowe· zostały· przetransformowane· na· wartości· dyskretne· porządkowe· w· oparciu· o· wartości· graniczne;· Poniżej· propozycja· pochodząca· od· eksperta· diagnostyki· samochodowej:¶
- **s1** → : → *(-∞,·74>,·(74,·79>,·(79,·85>,·(85,·∞)*¶
- **s2** → : → *(-∞,·2.2>,·(2.2,·2.4>,·(2.4,·∞)*¶
- **s3** → : → *(-∞,·59>,·(59,·∞)*¶
- **s4** → : → *(-∞,·441>,·(441,·∞)*¶
- **s5** → : → *(-∞,·22>,·(22,·∞)*¶
- **s6** → : → *(-∞,·25.2>,·(25.2,·∞)*¶
- **s7** → : → *(-∞,·1.2>,·(1.2,·∞)*¶
- **s8** → : → *(-∞,·119>,·(119,·∞)*¶

¶

Jeśli·to·potrzebne·możesz·rozważać·powyższą·propozycję,·ale·podstawową·analizę·proszę·wykonać·dla·danych· zdefiniowanych·na·skalach·numerycznych.¶

¶

Ogląd wczytanych danych

STATISTICA - autobusyplain.sta

File Edit View Insert Format Statistics Graphs Tools Data Window Help

Add to Workbook Add to Report

Arial 10 Vars Cases

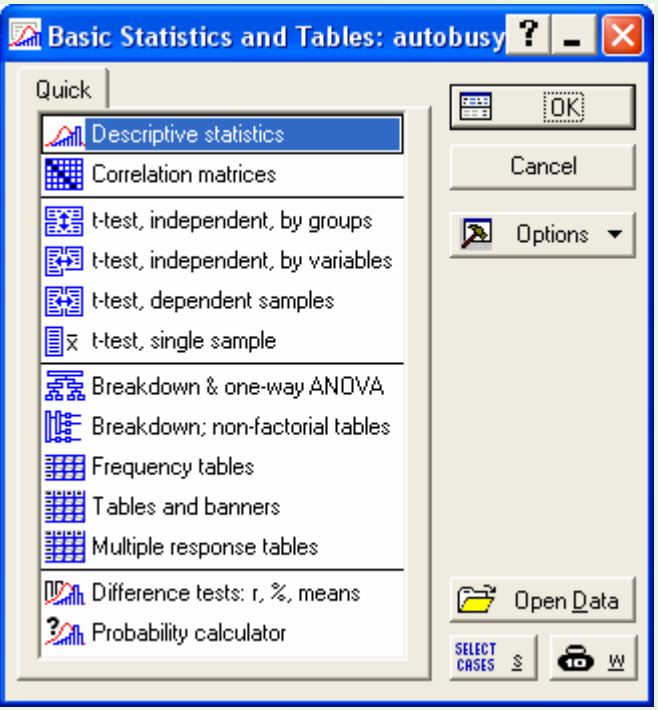
Data: autobusyplain.sta* (11v by 81c)

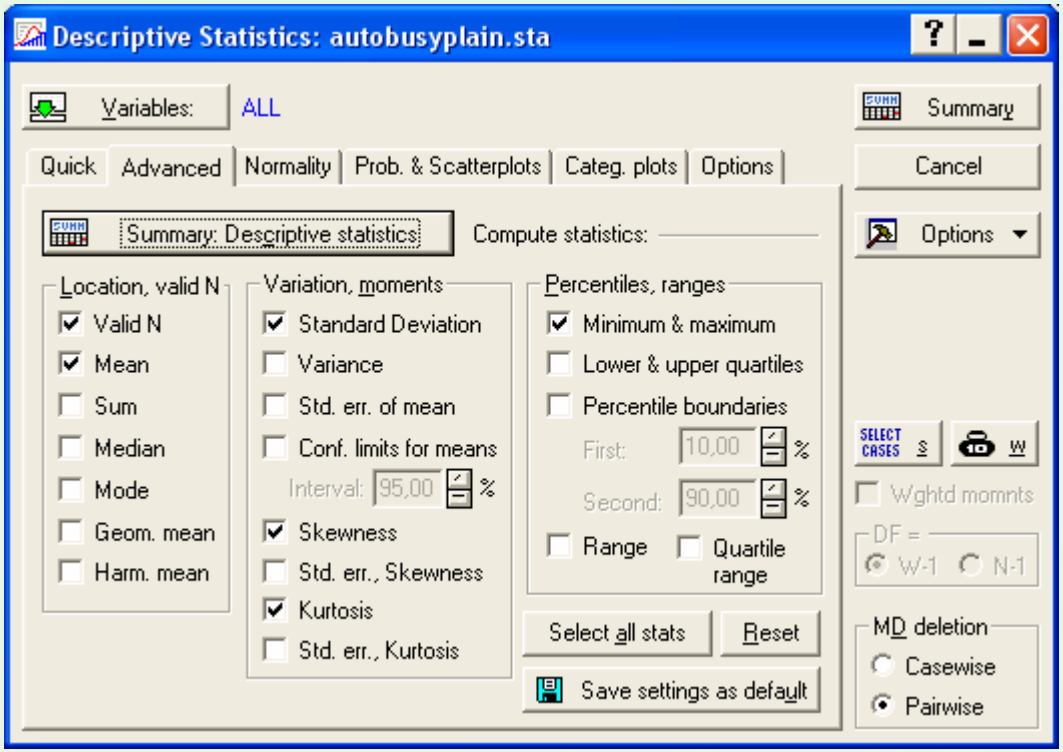
	1 id	2 MaxSpeed	3 Compr_preasure	4 blacking	5 torque	6 summer_cons	7 winter_cons	8 oil_cons	9 horsepower	10 D1	11 D2
1	1	90	2,52	38	481	21,8	26,4	0,7	145	1	1
2	2	76	2,11	70	420	22	25,5	2,7	110	2	3
3	3	63	1,96	82	400	22	24,8	3,7	101	2	3
4	4	90	2,48	49	477	21,9	25,1	1	138	1	1
5	5	85	2,45	52	460	21,8	25,2	1,4	130	1	2
6	6	72	2,2	73	425	23,1	27,4	2,8	112	2	3
7	7	88	2,5	50	480	21,6	24,7	1,1	140	1	1
8	8	87	2,48	56	465	22,8	27,6	1,4	135	1	1
9	9	90	2,56	16	486	26,5	27,3	0,2	150	1	1
10	10	60	1,95	95	400	23,3	24,8	4,4	96	2	3
11	11	80	2,41	60	451	21,7	26,1	1,7	125	1	2
12	12	78	2,4	63	448	21,8	26	1,9	120	1	2
13	13	90	2,58	26	482	22,4	24,5	0,4	148	1	1
14	14	62	1,98	93	400	22	28,4	3,9	100	2	3
15	15	82	2,5	54	461	22	26,3	1,4	132	1	2
16	16	65	2,22	67	402	22	23,9	2,3	103	2	3
17	17	90	2,48	51	468	22	26,5	1,2	138	1	1
18	18	90	2,6	15	488	20	23,2	0,1	150	1	1
19	19	76	2,39	65	428	27	33,4	2	116	2	3
20	20	85	2,42	50	454	21,5	26,3	1,3	129	1	2
21	21	85	2,41	58	450	22	25,5	1,5	126	1	2
22	22	88	2,47	48	458	22,4	25,1	1,1	130	1	1
23	23	60	1,93	90	400	24	28,7	4	95	2	3
24	24	64	2,2	71	420	23,1	25,2	2,6	105	2	3
25	25	75	2,39	64	432	22,2	25,1	1,7	114	2	2
26	26	74	2,36	64	420	21,9	25,4	1,9	110	2	2
27	27	68	2,15	70	400	22	26	2,6	100	2	3

Descriptive Statistics:...

Ocena jakości – statystyki podstawowe

- Statistica – descriptive statistics

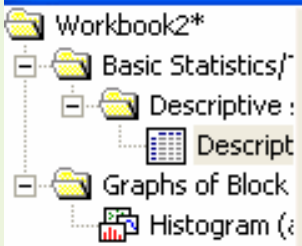




Raport podstawowych statystyk opisowych

- Przeanalizujmy podstawowe miary

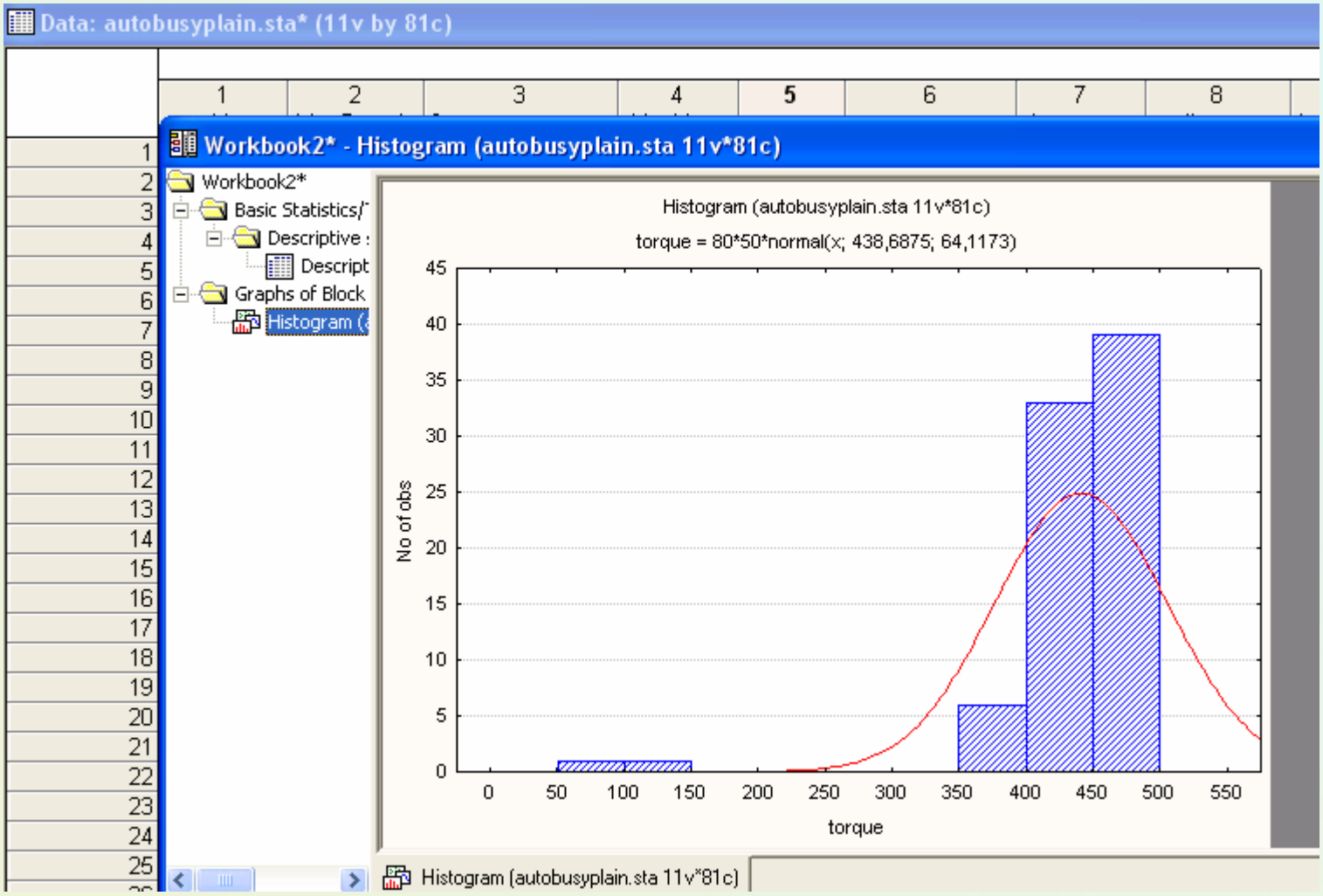
Workbook2* - Descriptive Statistics (autobusyplain.sta)



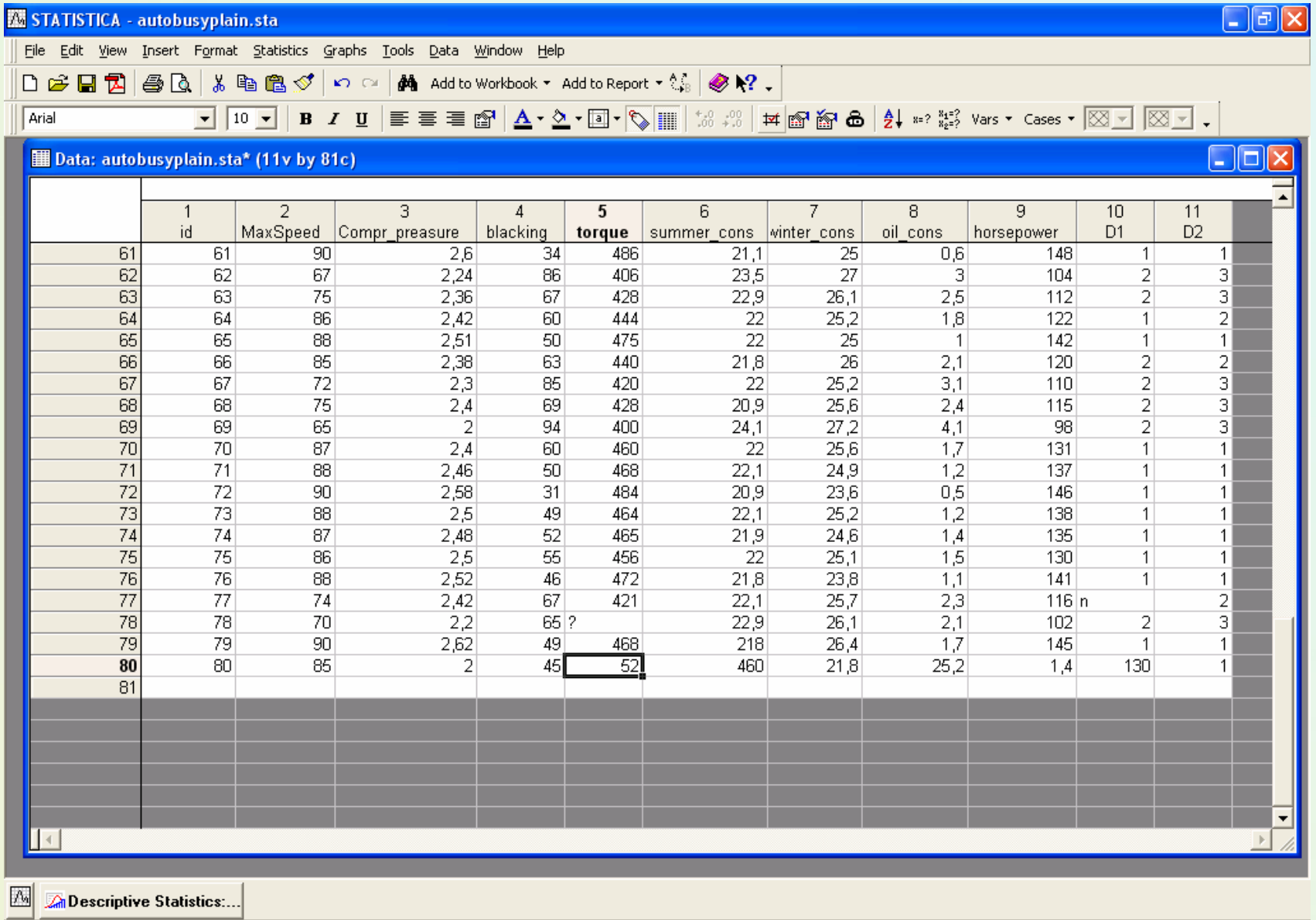
Variable	Descriptive Statistics (autobusyplain.sta)						
	Valid N	Mean	Minimum	Maximum	Std.Dev.	Skewness	Kurtosis
id	80	40,5000	1,00000	80,0000	23,23790	0,00000	-1,20000
MaxSpeed	80	81,0625	60,00000	90,0000	8,91357	-0,85985	-0,51579
Compr_preasure	80	2,3909	1,93000	2,6200	0,16375	-1,28523	1,19673
blacking	80	57,9750	15,00000	95,0000	16,67179	-0,08102	0,61765
torque	80	438,6875	52,00000	488,0000	64,11730	-4,68915	25,52812
summer_cons	80	30,1650	20,00000	460,0000	53,36880	7,34541	56,44600
winter_cons	80	25,5763	21,80000	33,4000	1,38292	2,24505	12,78355
oil_cons	80	2,0775	0,10000	25,2000	2,77165	7,52863	63,07514
horsepower	80	123,1800	1,40000	150,0000	20,58264	-2,60904	14,24515
D1	80	4,2500	1,00000	130,0000	18,08104	6,35214	40,03185
D2	80	1,8750	1,00000	3,0000	0,87692	0,24908	-1,66806

Spójrzmy dokładniej na wybrane atrybuty

- Np. moment obrotowy (porównaj z definicją dziedziny)



Poszukiwanie źródła trudności



	1 id	2 MaxSpeed	3 Compr_preasure	4 blacking	5 **torque**	6 summer_cons	7 winter_cons	8 oil_cons	9 horsepower	10 D1	11 D2
61	61	90	2,6	34	486	21,1	25	0,6	148	1	1
62	62	67	2,24	86	406	23,5	27	3	104	2	3
63	63	75	2,36	67	428	22,9	26,1	2,5	112	2	3
64	64	86	2,42	60	444	22	25,2	1,8	122	1	2
65	65	88	2,51	50	475	22	25	1	142	1	1
66	66	85	2,38	63	440	21,8	26	2,1	120	2	2
67	67	72	2,3	85	420	22	25,2	3,1	110	2	3
68	68	75	2,4	69	428	20,9	25,6	2,4	115	2	3
69	69	65	2	94	400	24,1	27,2	4,1	98	2	3
70	70	87	2,4	60	460	22	25,6	1,7	131	1	1
71	71	88	2,46	50	468	22,1	24,9	1,2	137	1	1
72	72	90	2,58	31	484	20,9	23,6	0,5	146	1	1
73	73	88	2,5	49	464	22,1	25,2	1,2	138	1	1
74	74	87	2,48	52	465	21,9	24,6	1,4	135	1	1
75	75	86	2,5	55	456	22	25,1	1,5	130	1	1
76	76	88	2,52	46	472	21,8	23,8	1,1	141	1	1
77	77	74	2,42	67	421	22,1	25,7	2,3	116	n	2
78	78	70	2,2	65	?	22,9	26,1	2,1	102	2	3
79	79	90	2,62	49	468	218	26,4	1,7	145	1	1
80	80	85	2	45	52	460	21,8	25,2	1,4	130	1
81											

Spójrz do oryginalnego pliku

autobusy.isf - Notatnik

Plik Edycja Format Widok Pomoc

```
51, 90,   2.54,   40,   480,   21.8,   25.1,   0.9,   145,    1,   1
52, 89,   2.51,   46,   470,   22  ,   25.5,   1.1,   141,    1,   1
53, 85,   2.48,   60,   440,   22.2,   26.3,   1.8,   120,    1,   2
54, 90,   2.52,   45,   479,   21.6,   25.1,   1  ,   145,    1,   1
55, 90,   2.54,   47,   481,   22  ,   24.9,   1  ,   145,    1,   1
56, 85,   2.5 ,   60,   446,   21.9,   24.6,   1.9,   123,    1,   1
57, 88,   2.56,   50,   465,   21.6,   24.5,   1.2,   137,    1,   1
58, 84,   2.32,   65,   428,   22.0,   25.1,   2.2,   116,    2,   3
59, 86,   2.47,   54,   452,   22  ,   25.1,   1.6,   125,    1,   1
60, 66,   2.29,   86,   402,   23.3,   26.6,   3.2,   105,    2,   3
61, 90,   2.60,   34,   486,   21.1,   25  ,   0.6,   148,    1,   1
62, 67,   2.24,   86,   406,   23.5,   27  ,   3  ,   104,    2,   3
63, 75,   2.36,   67,   428,   22.9,   26.1,   2.5,   112,    2,   3
64, 86,   2.42,   60,   444,   22  ,   25.2,   1.8,   122,    1,   2
65, 88,   2.51,   50,   475,   22  ,   25  ,   1  ,   142,    1,   1
66, 85,   2.38,   63,   440,   21.8,   26  ,   2.1,   120,    2,   2
67, 72,   2.30,   85,   420,   22  ,   25.2,   3.1,   110,    2,   3
68, 75,   2.4 ,   69,   428,   20.9,   25.6,   2.4,   115,    2,   3
69, 65,   2   ,   94,   400,   24.1,   27.2,   4.1,   98 ,    2,   3
70, 87,   2.4 ,   60,   460,   22  ,   25.6,   1.7,   131,    1,   1
71, 88,   2.46,   50,   468,   22.1,   24.9,   1.2,   137,    1,   1
72, 90,   2.58,   31,   484,   20.9,   23.6,   0.5,   146,    1,   1
73, 88,   2.5 ,   49,   464,   22.1,   25.2,   1.2,   138,    1,   1
74, 87,   2.48,   52,   465,   21.9,   24.6,   1.4,   135,    1,   1
75, 86,   2.5 ,   55,   456,   22.0,   25.1,   1.5,   130,    1,   1
76, 88,   2.52,   46,   472,   21.8,   23.8,   1.1,   141,    1,   1
77, 74,   2.42,   67,   421,   22.1,   25.7,   2.3,   116,    n,   2
78, 70,   2.2,    65,   ?     22.9,   26.1,   2.1,   102,    2,   3
79, 90,   2.62,   49,   468,    218,   26.4,   1.7,   145,    1,   1
80, 85,   2,45    52,   460,   21.8,   25.2,   1.4,   130,    1,   2
**END
```

Poprawki w pliku

- Przesunąć wpis wiersza id. 80
- zapisy ? – wartości średnie w klasie decyzyjnej
- Symbol spoza dziedziny „n”
  - pewnie niesprawne autobusy – kod 2

Zmiany statystyk opisowych

Workbook2* - Descriptive Statistics (autobusyplainpopraw.sta)

atistics/Tables (autobus
criptive statistics dialog
Descriptive Statistics (au
Descriptive Statistics (au
of Block Data (autobusy
ogram (autobusyplain.st

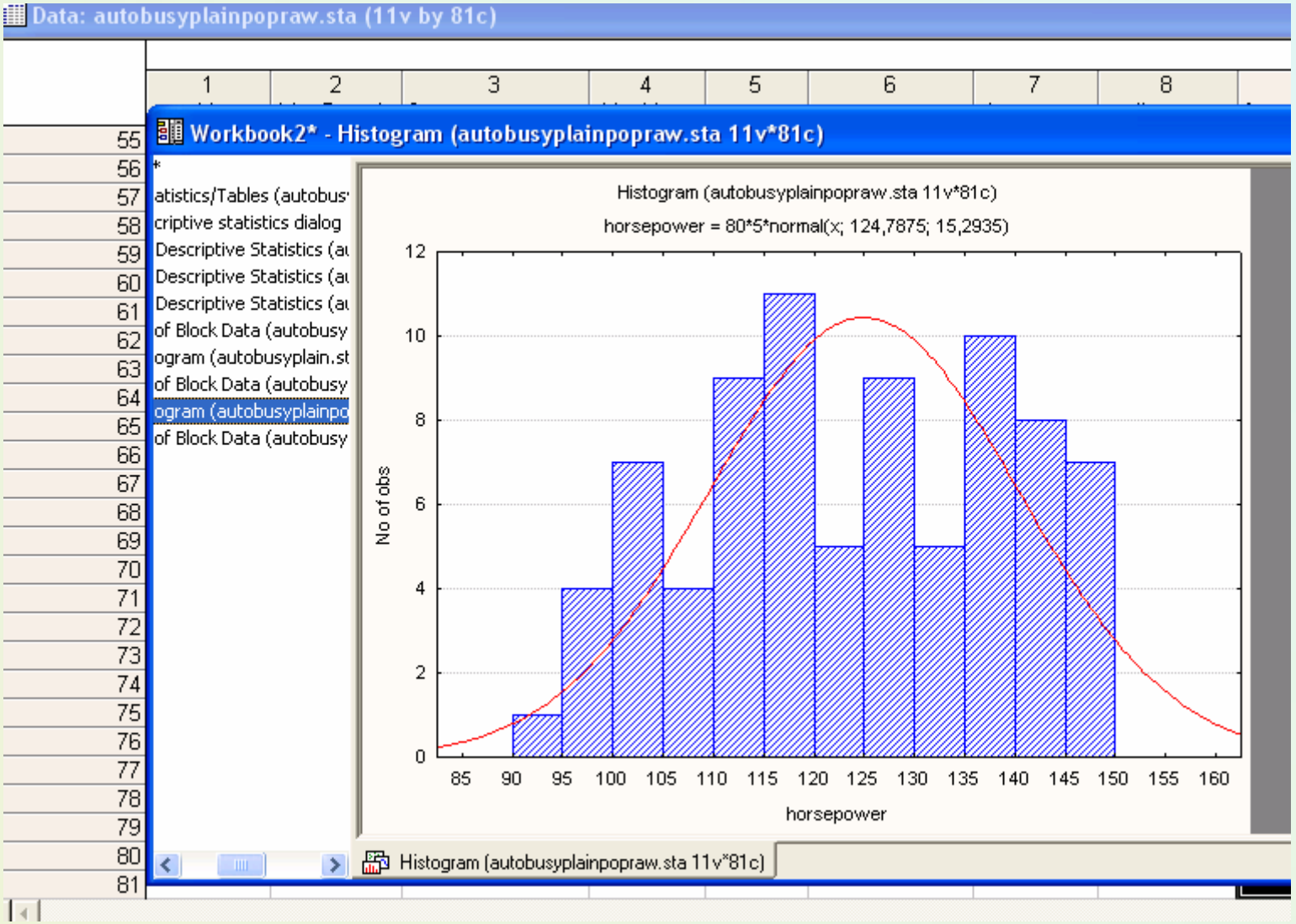
Descriptive Statistics (autobusyplainpopraw.sta)

Variable	Valid N	Mean	Minimum	**Maximum**	Std.Dev.	Skewness	Kurtosis
id	80	40,5000	1,0000	80,0000	23,23790	0,00000	-1,20000
MaxSpeed	80	81,0625	60,0000	90,0000	8,91357	-0,85985	-0,51579
Compr_preasure	80	2,3965	1,9300	2,6200	0,15777	-1,34854	1,56840
blacking	80	58,0375	15,0000	95,0000	16,63186	-0,08844	0,65062
torque	80	447,9334	400,0000	488,0000	26,35268	-0,31996	-1,00724
summer_cons	80	24,6875	20,0000	218,0000	21,90940	8,91586	79,65439
winter_cons	80	25,6187	23,2000	33,4000	1,31603	2,81213	14,95687
oil_cons	80	1,7800	0,1000	4,4000	0,91131	0,76994	0,60478
horsepower	80	124,7875	95,0000	150,0000	15,29349	-0,12852	-1,07744
D1	80	1,4000	1,0000	2,0000	0,49299	0,41609	-1,87438
D2	80	1,8875	1,0000	3,0000	0,87140	0,22289	-1,65965

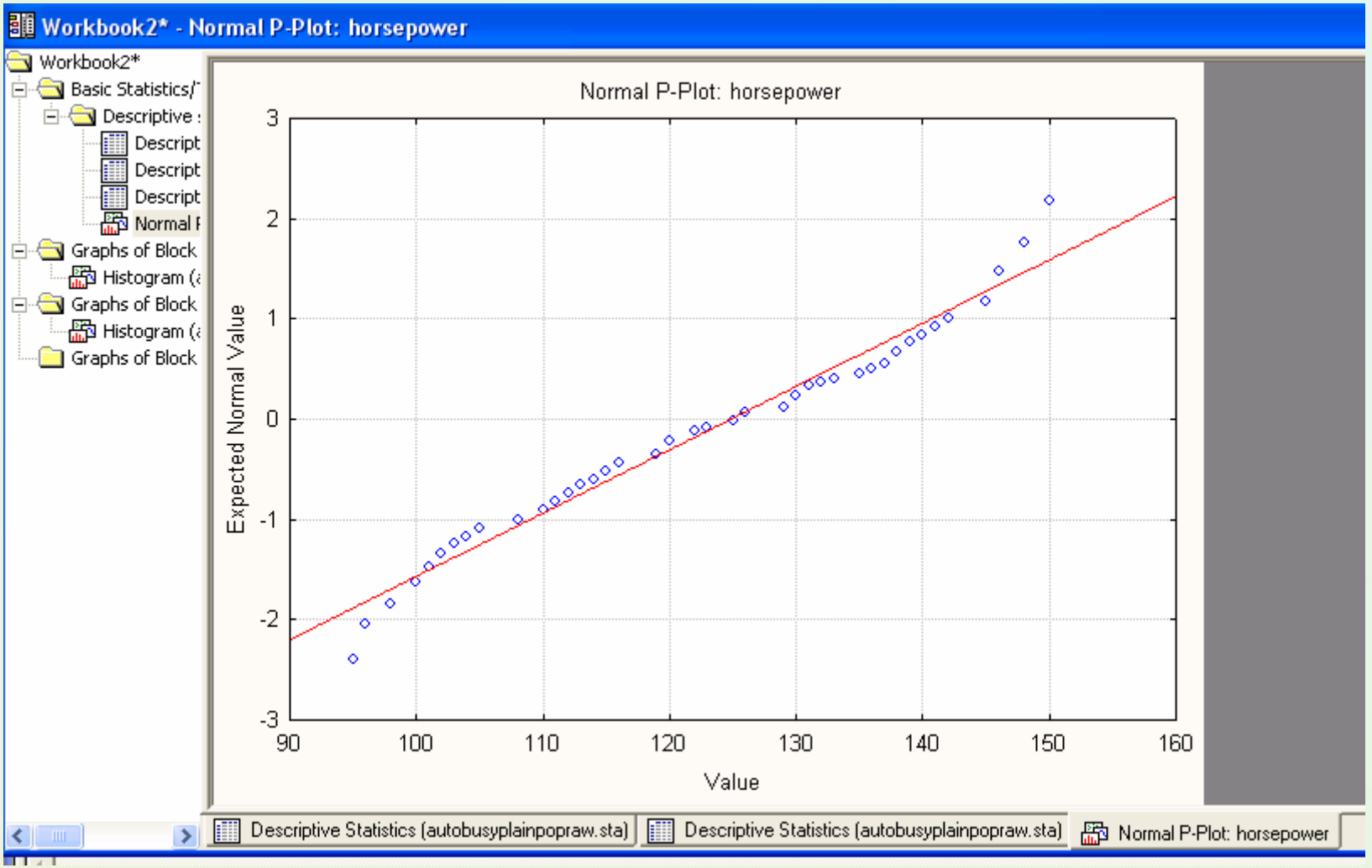
Descriptive Statistics (autobusyplain.sta) | Descriptive Statistics (autobusyplainpopraw.sta)

- 218? Problem jest z separatorem „ ,”
- 218 → 21,8

Sprawdzenie mocy silnika (horsepower)



Spojrzenie na rozkłady



Analiza współzależności

- Macierz korelacji

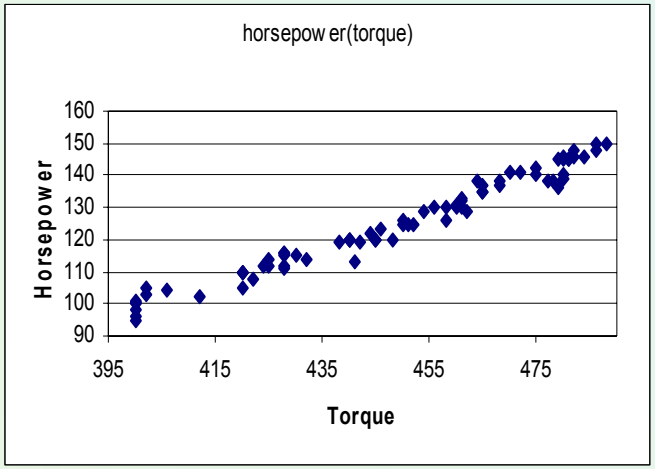
elations (autobusyplainpopraw.sta)

Correlations (autobusyplainpopraw.sta)
Marked correlations are significant at p < ,05000
N=80 (Casewise deletion of missing data)

ariable	**MaxSpeed**	Compr_preasure	blacking	torque	summer_cons	winter_cons	oil_cons	horsepower	D1	D2
axSpeed	1,00	0,90	-0,85	0,93	-0,37	-0,34	-0,89	0,93	-0,84	-0,86
ompr_preasure	0,90	1,00	-0,85	0,88	-0,36	-0,34	-0,91	0,89	-0,75	-0,80
acking	-0,85	-0,85	1,00	-0,90	0,33	0,40	0,96	-0,91	0,73	0,80
rque	0,93	0,88	-0,90	1,00	-0,37	-0,37	-0,92	0,98	-0,85	-0,89
mmer_cons	-0,37	-0,36	0,33	-0,37	1,00	0,74	0,35	-0,37	0,38	0,39
nter_cons	-0,34	-0,34	0,40	-0,37	0,74	1,00	0,38	-0,36	0,34	0,40
_cons	-0,89	-0,91	0,96	-0,92	0,35	0,38	1,00	-0,91	0,77	0,82
rsepower	0,93	0,89	-0,91	0,98	-0,37	-0,36	-0,91	1,00	-0,84	-0,90

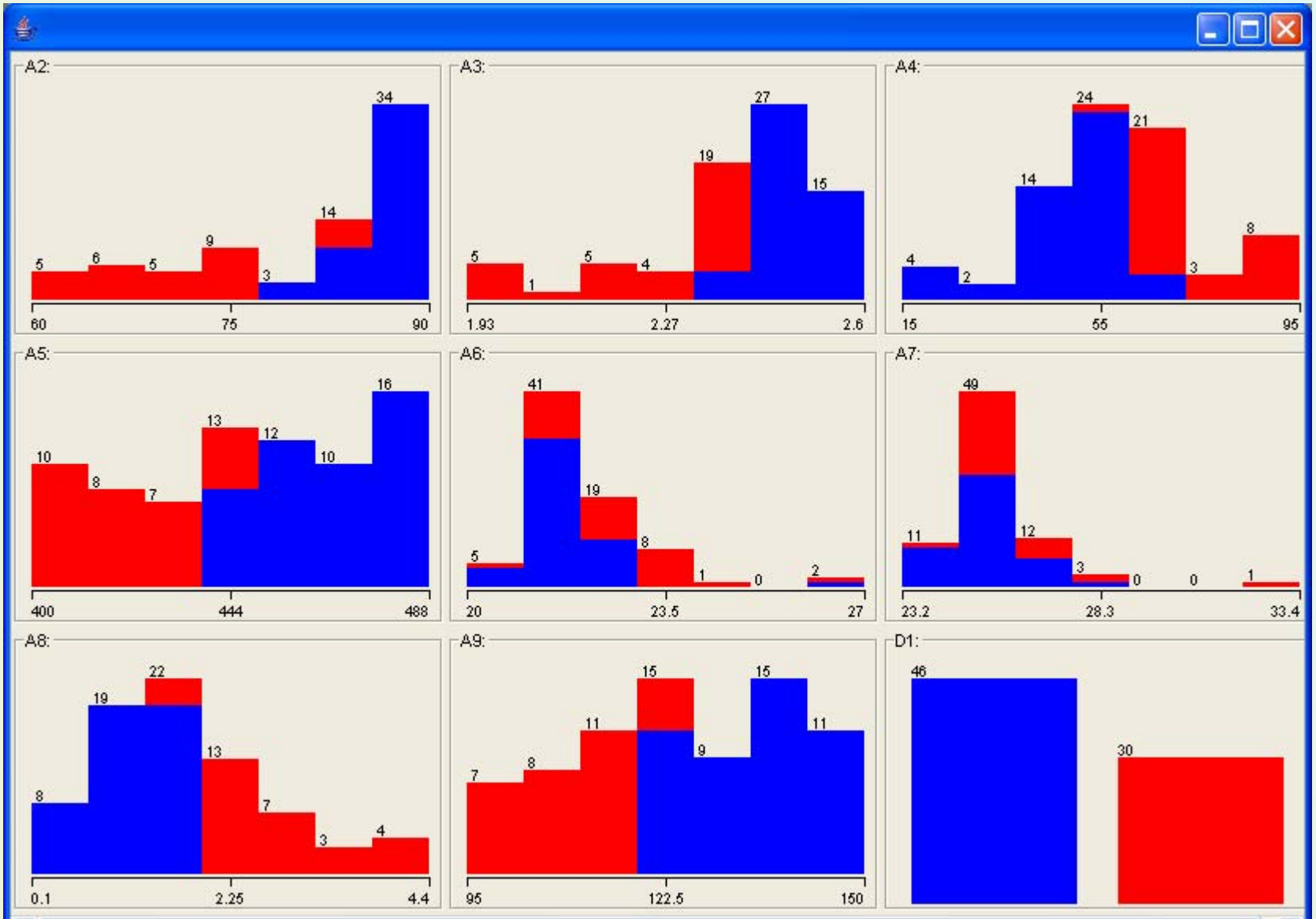
Powiązanie pewnych symptomów

- Można zrobić wykresy korelacyjne (rozrzutu XY)

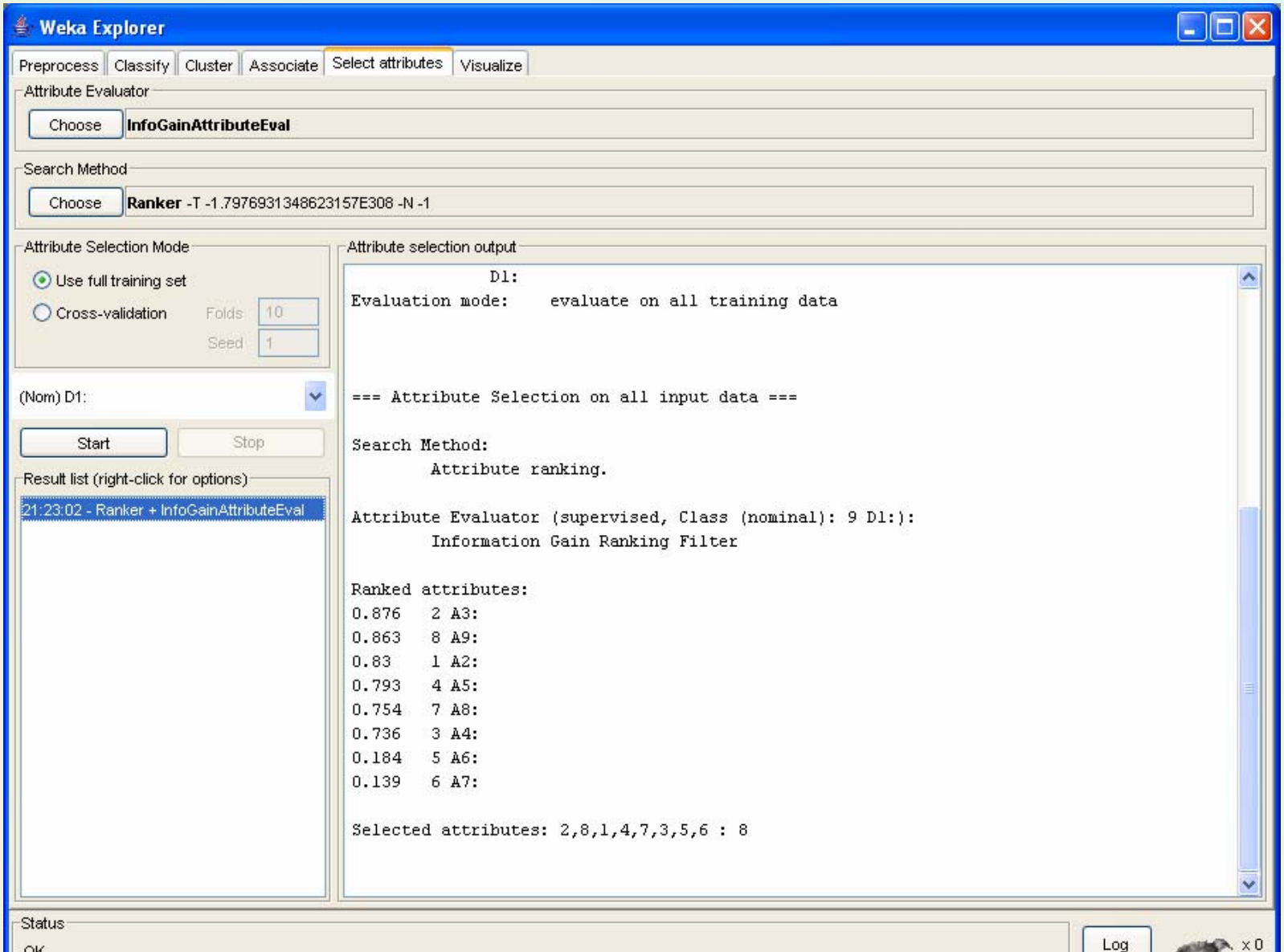


- Wiedza dziedzinowa
  - ciśnienie sprężania powiązane jest z momentem obrotowym, im większe ciśnienie tym większy można uzyskać moment obrotowy,
  - moment obrotowy powiązany jest z mocą pojazdu,
  - zawartość elementów smołowatych i zużycie oleju świadczyć może o wieku silnika i jego stanie technicznym,
  - Mniejsza użyteczność info. o zużyciu paliwa (warunki, styl jazdy,..)

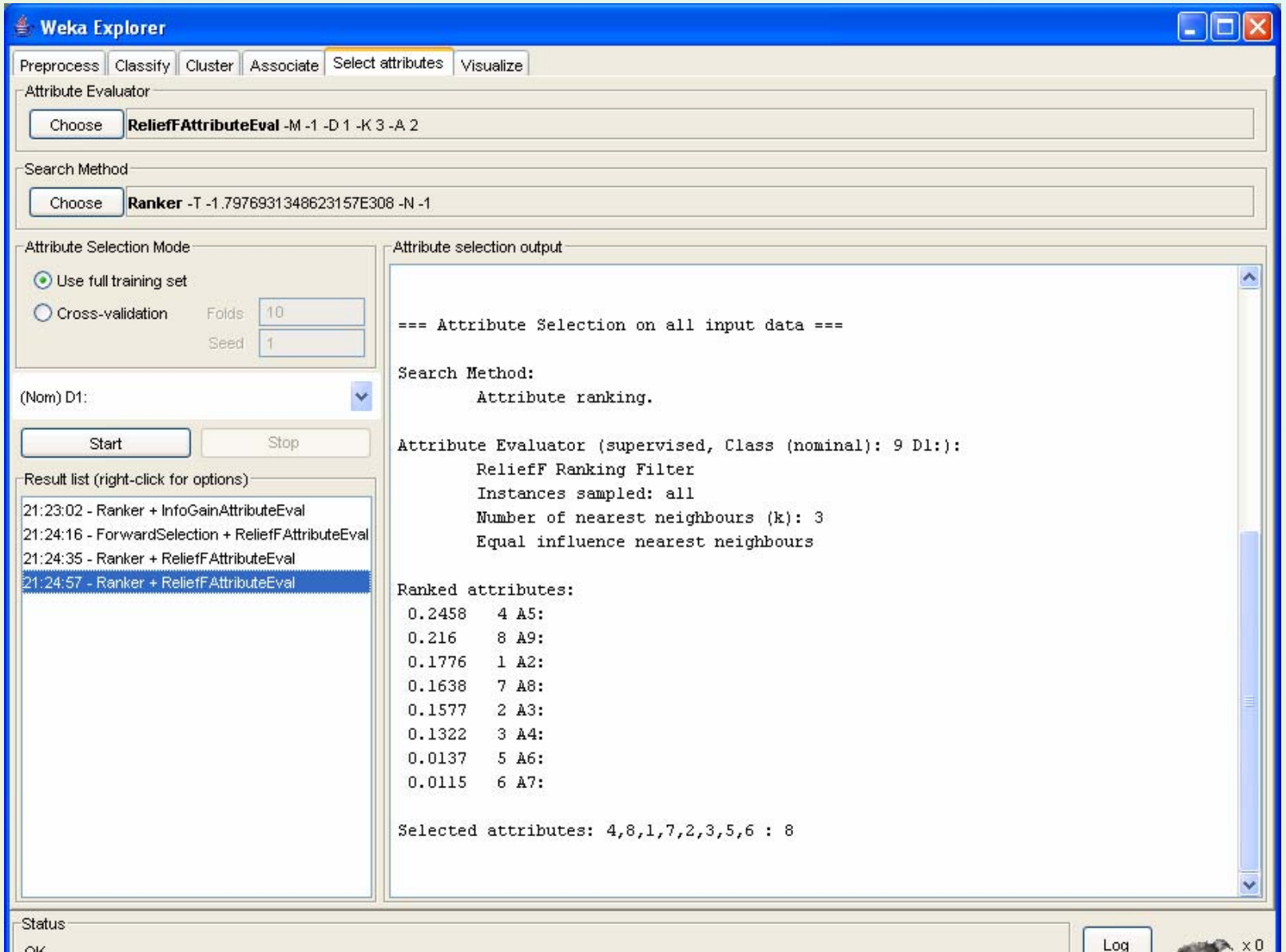
WEKA –visualize



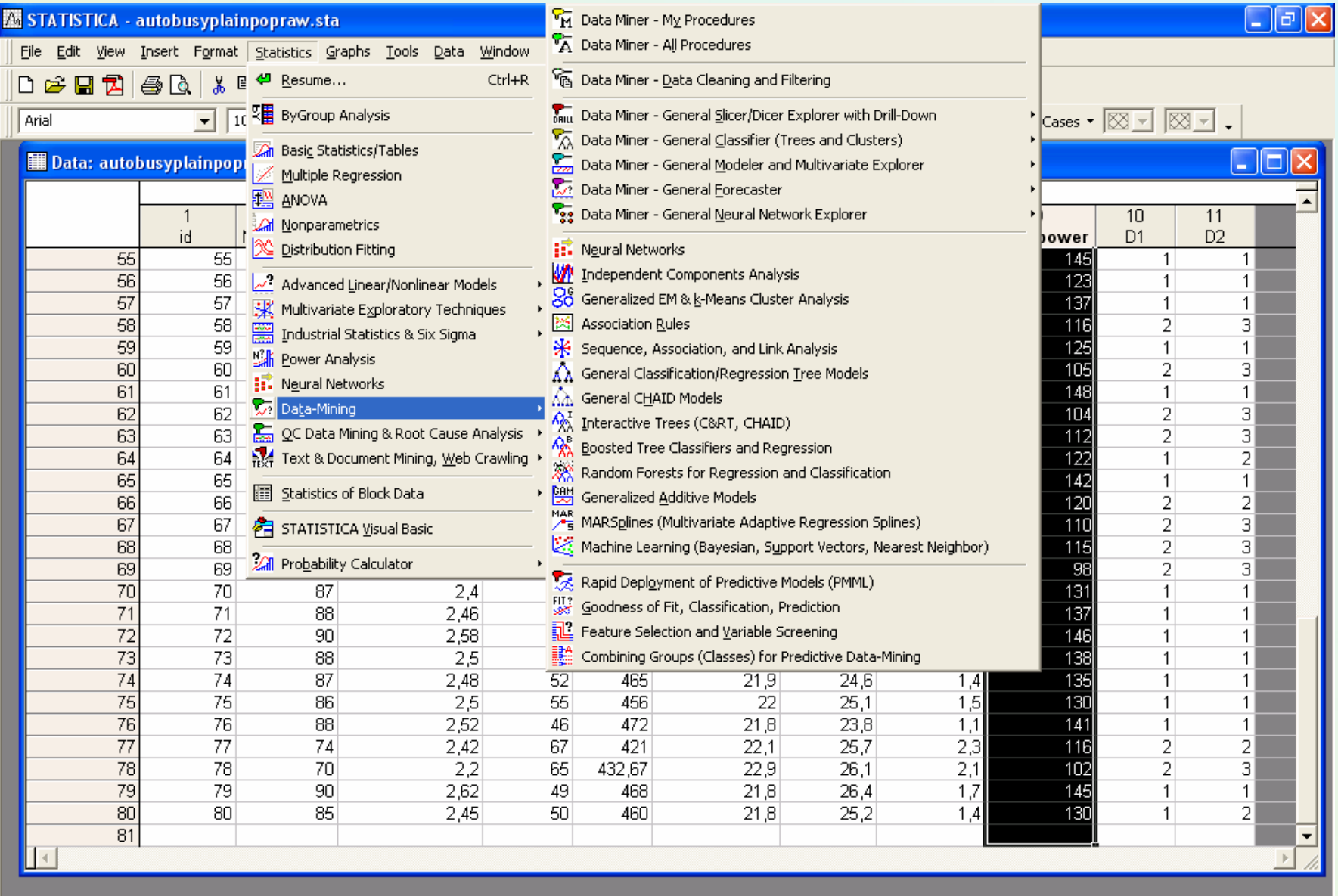
Select attributes z WEKA



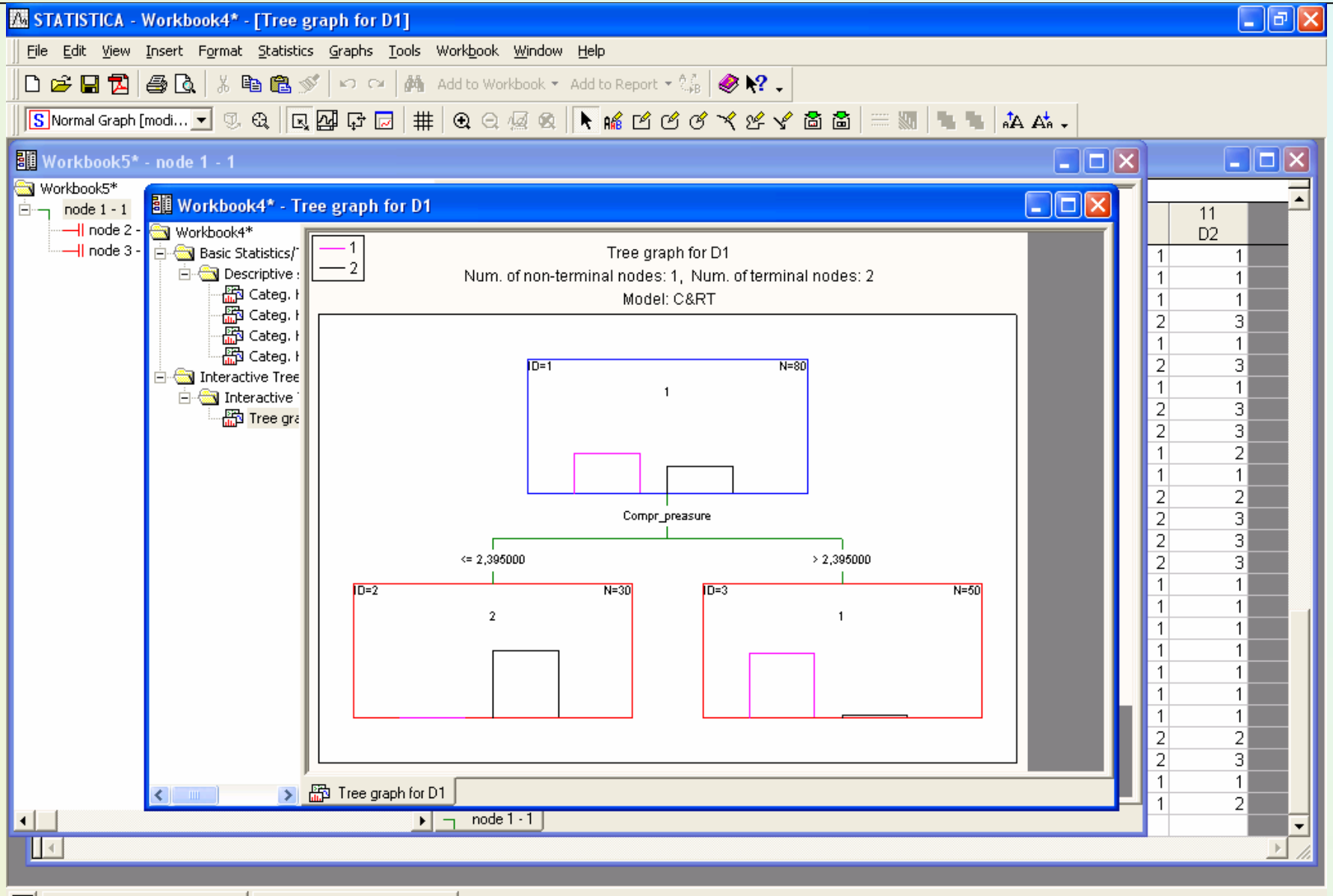
Relief (ważność z wagowaniem k-NN)



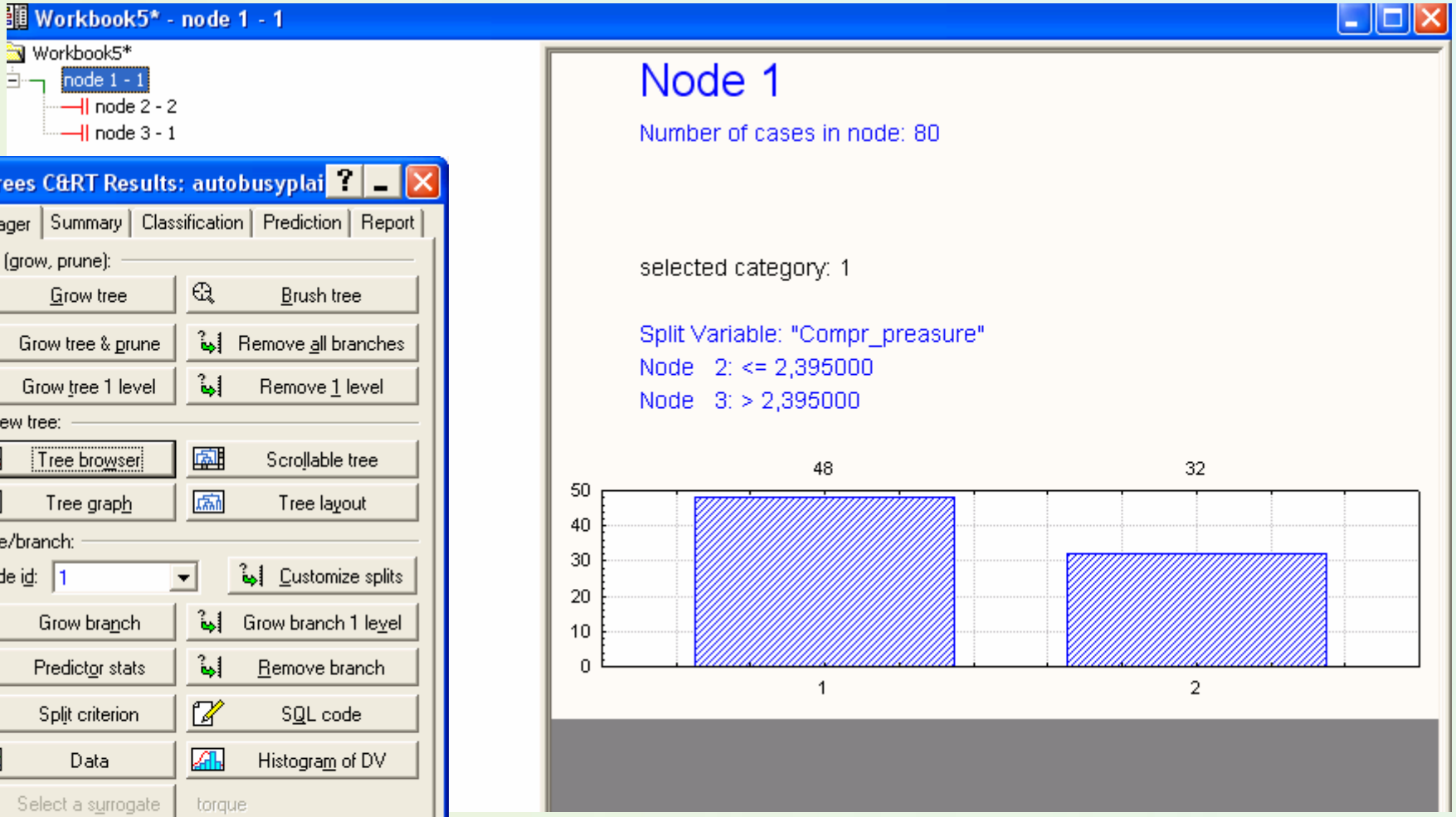
Statistica – Data Miner



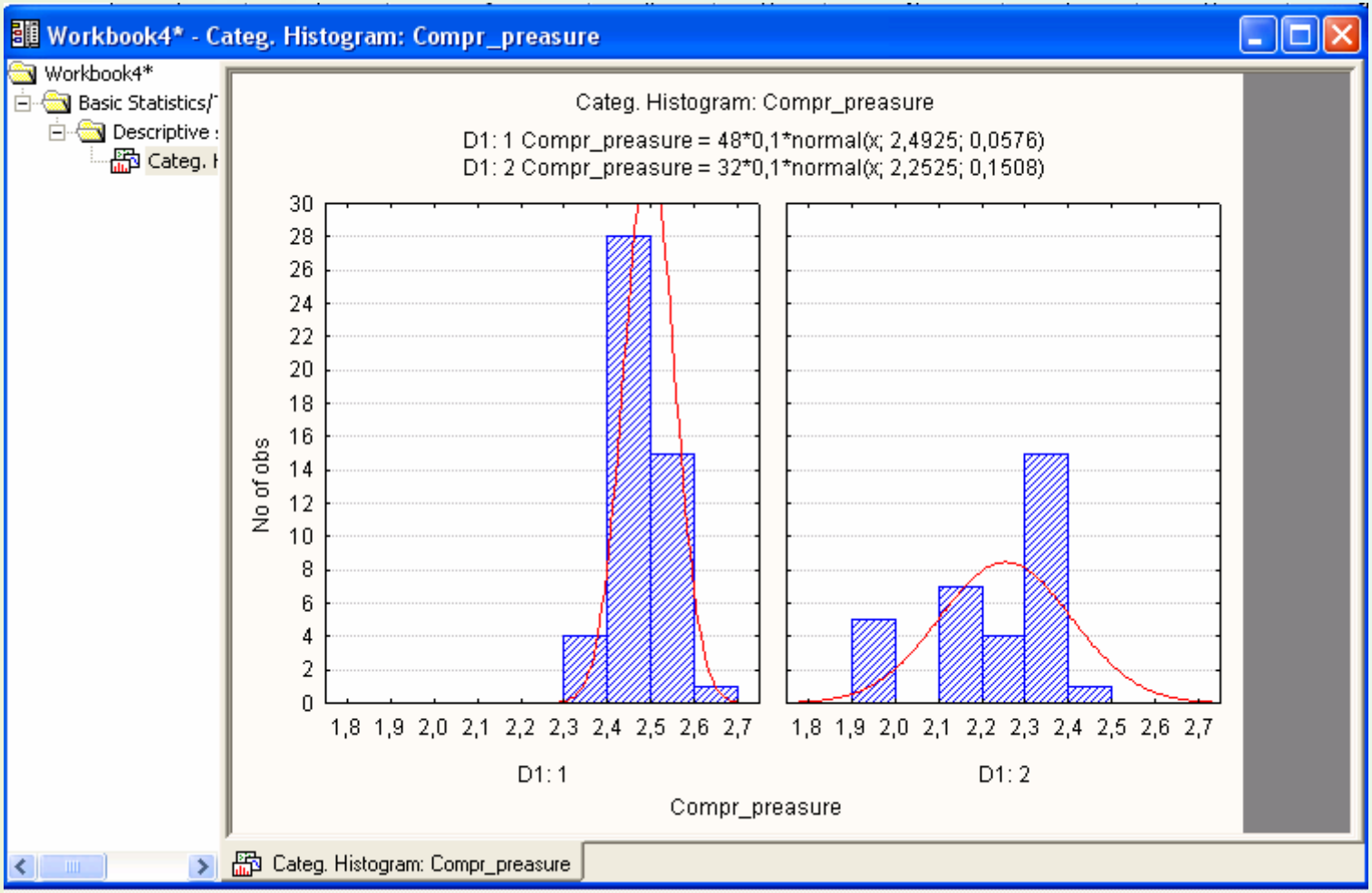
CART - tree


STATISTICA - Workbook4* - [Tree graph for D1]
File Edit View Insert Format Statistics Graphs Tools Workbook Window Help
Add to Workbook
Add to Report
Normal Graph [modi...
Workbook5* - node 1 - 1
Workbook5*
node 1 - 1
node 2 -
node 3 -
Workbook4* - Tree graph for D1
Workbook4*
Basic Statistics/
Descriptive
Categ.
Categ.
Categ.
Categ.
Interactive Tree
Interactive
Tree gra
1
2
Tree graph for D1
Num. of non-terminal nodes: 1, Num. of terminal nodes: 2
Model: C&RT
ID=1
N=80
1
Compr_preasure
<= 2,395000
> 2,395000
ID=2
N=30
2
ID=3
N=50
1
Tree graph for D1
node 1 - 1
11
D2

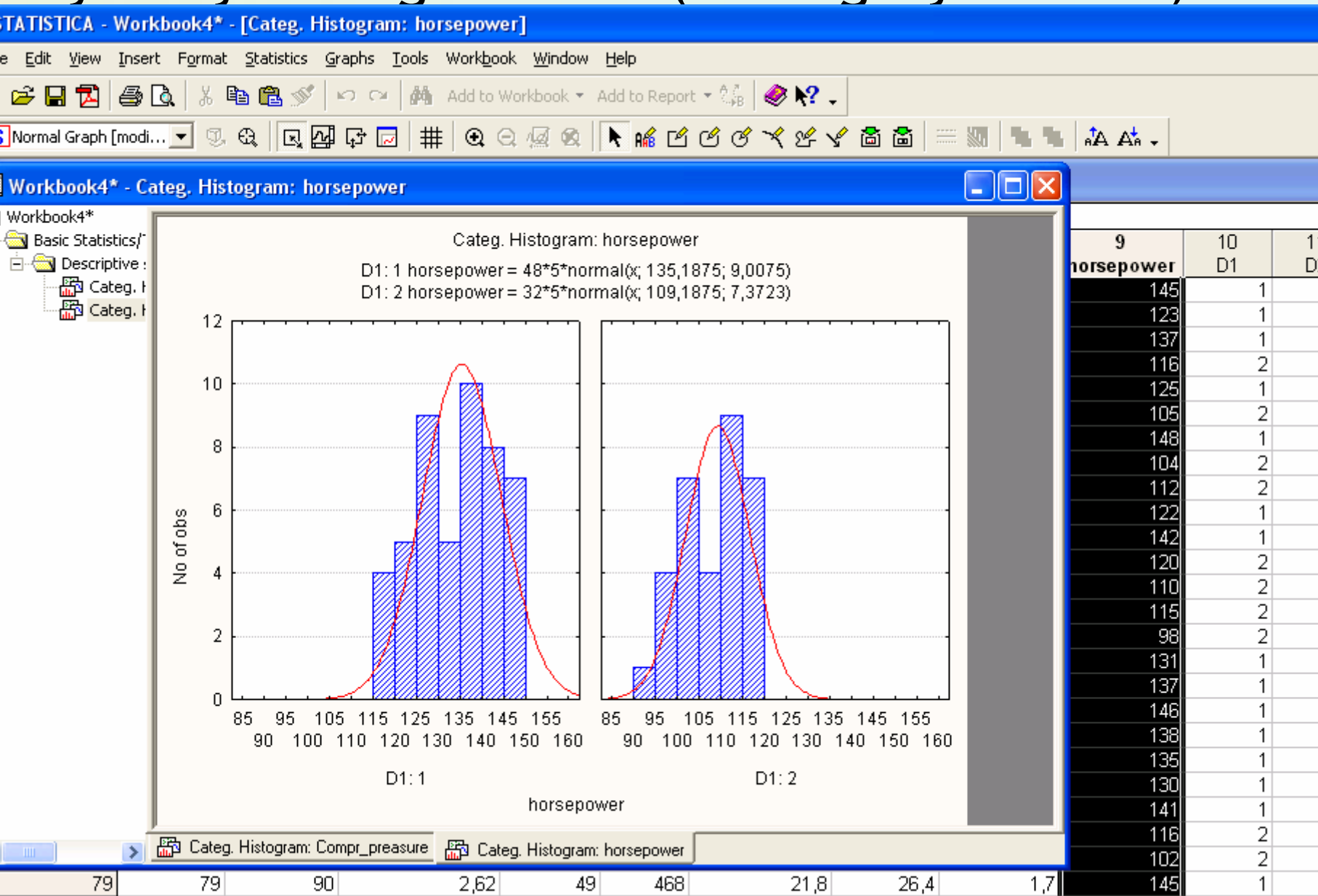
Przeglądanie węzłów drzewa (dec 1)



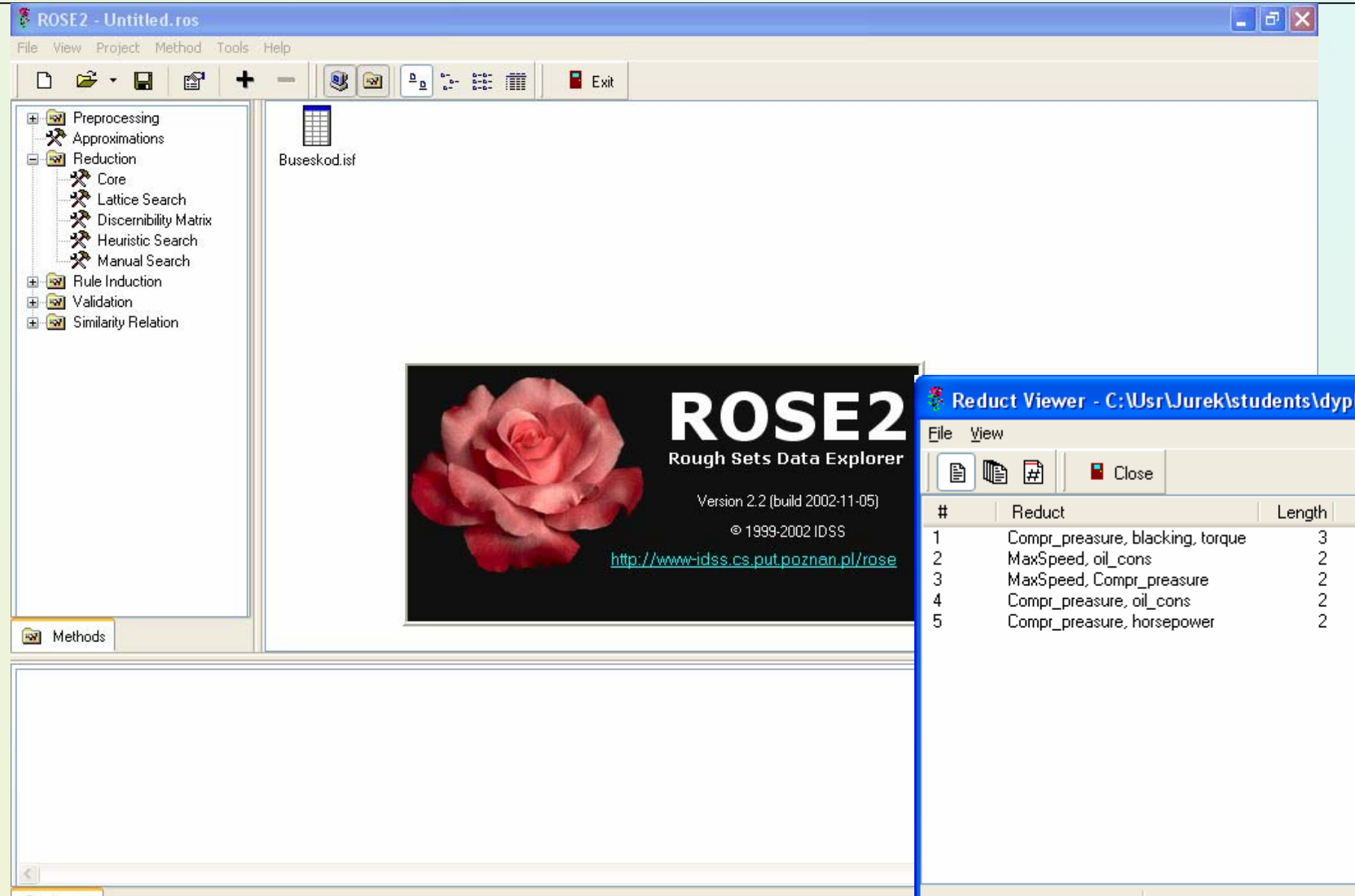
Skategoryzowane histogramy



Wykresy histogramowe (skategoryzowane)



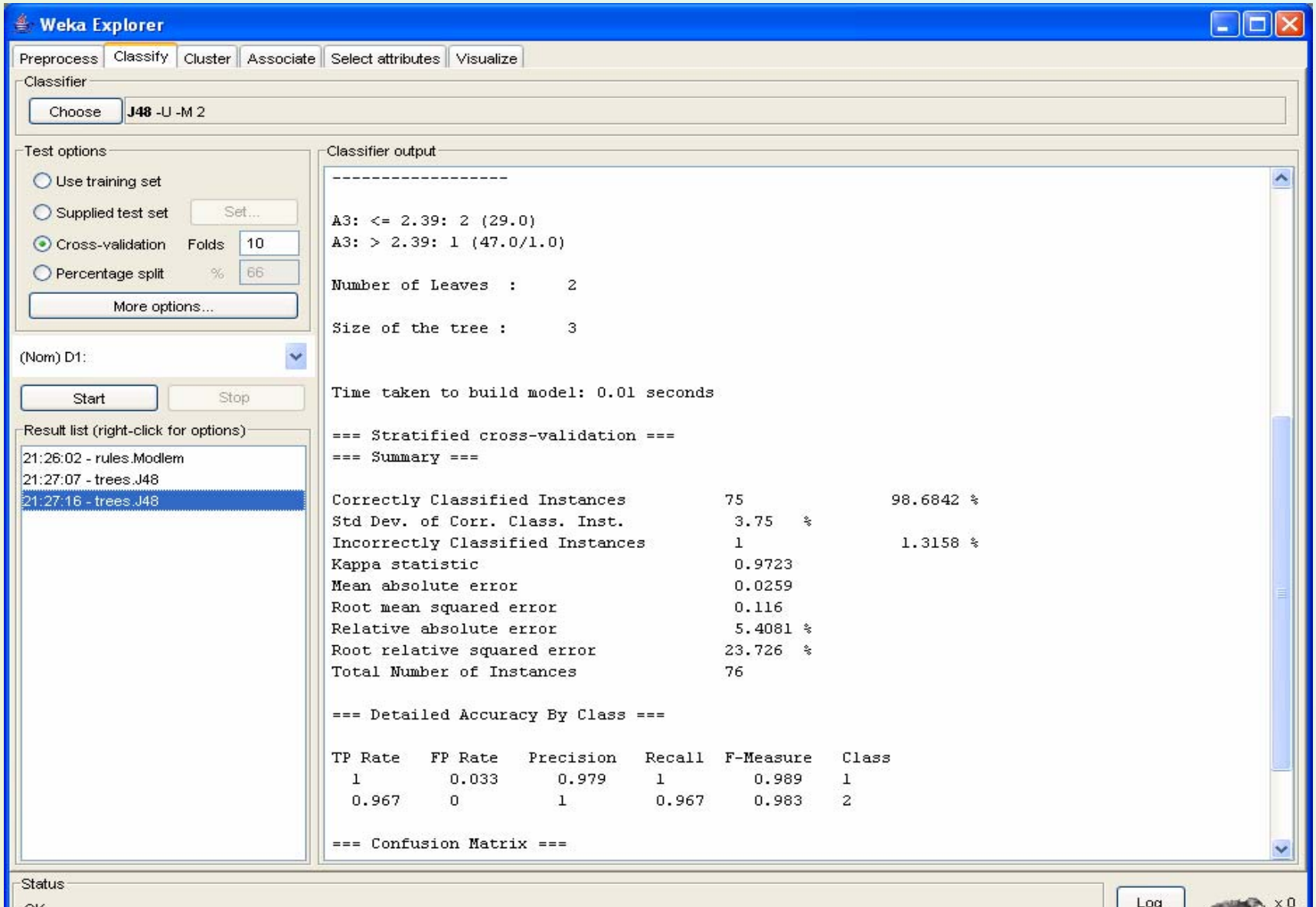
Podejście teorii zbiorów przybliżonych



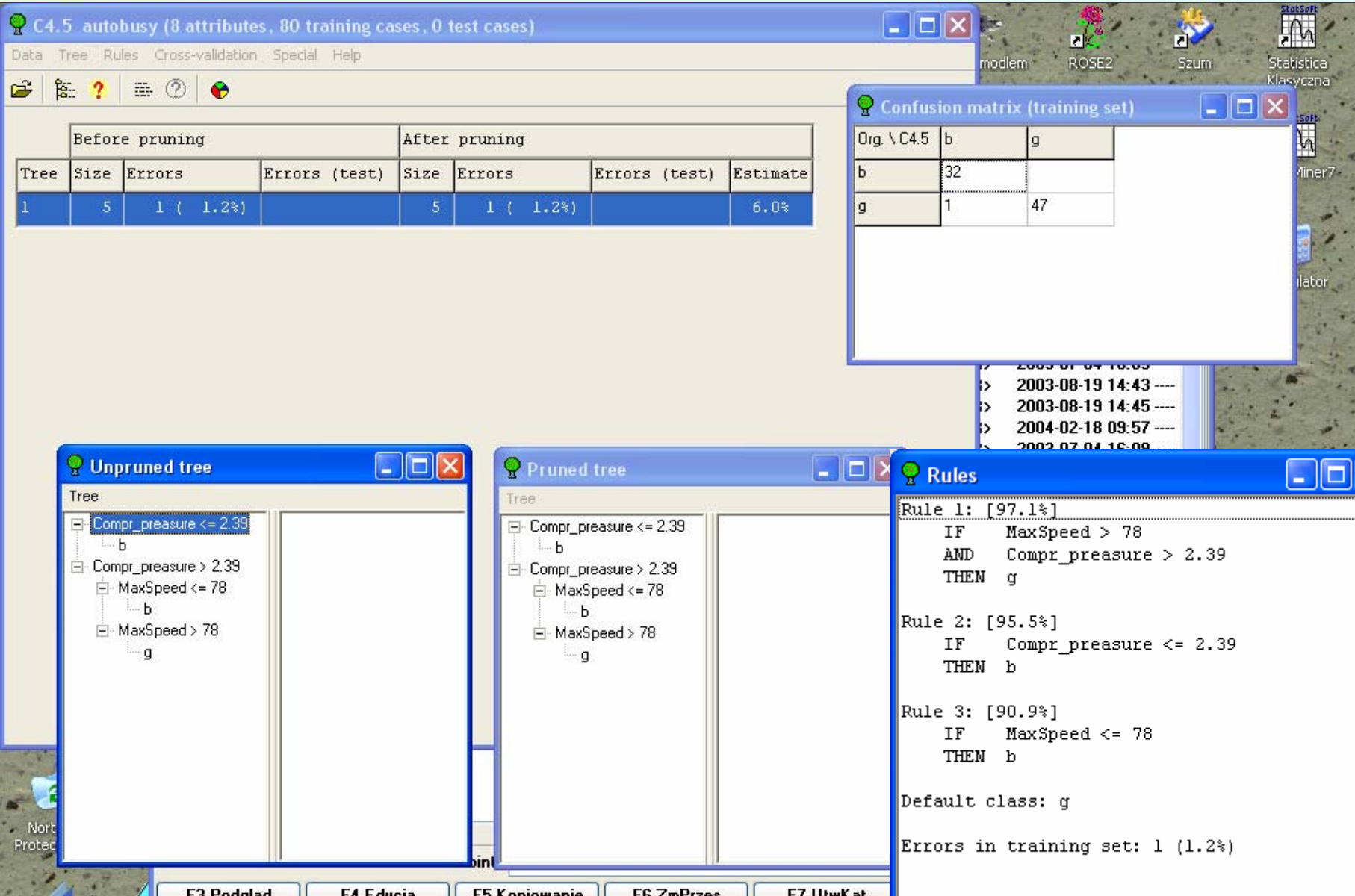
Dyskretne dane + teoria zbiorów przybliżonych
Hierarchia ważności symptomów dla klasyfikacji autobusów

- Można ustalać względna ważność symptomów dla klasyfikacji przykładów - przykładowa hierarchia ważności s8, s1, s3, s4 inne mniej istotne - analiza statystyczna
- w przypadku dyskretyzacji dziedzin wartości w oparciu o normy możliwe ustalenie podzbiorów zredukowanych {s2,s3,s8}, {s2,s3,s4}, {s1,s2} - teoria zbiorów przybliżonych
- Notacja s1 – prędkość maksymalna

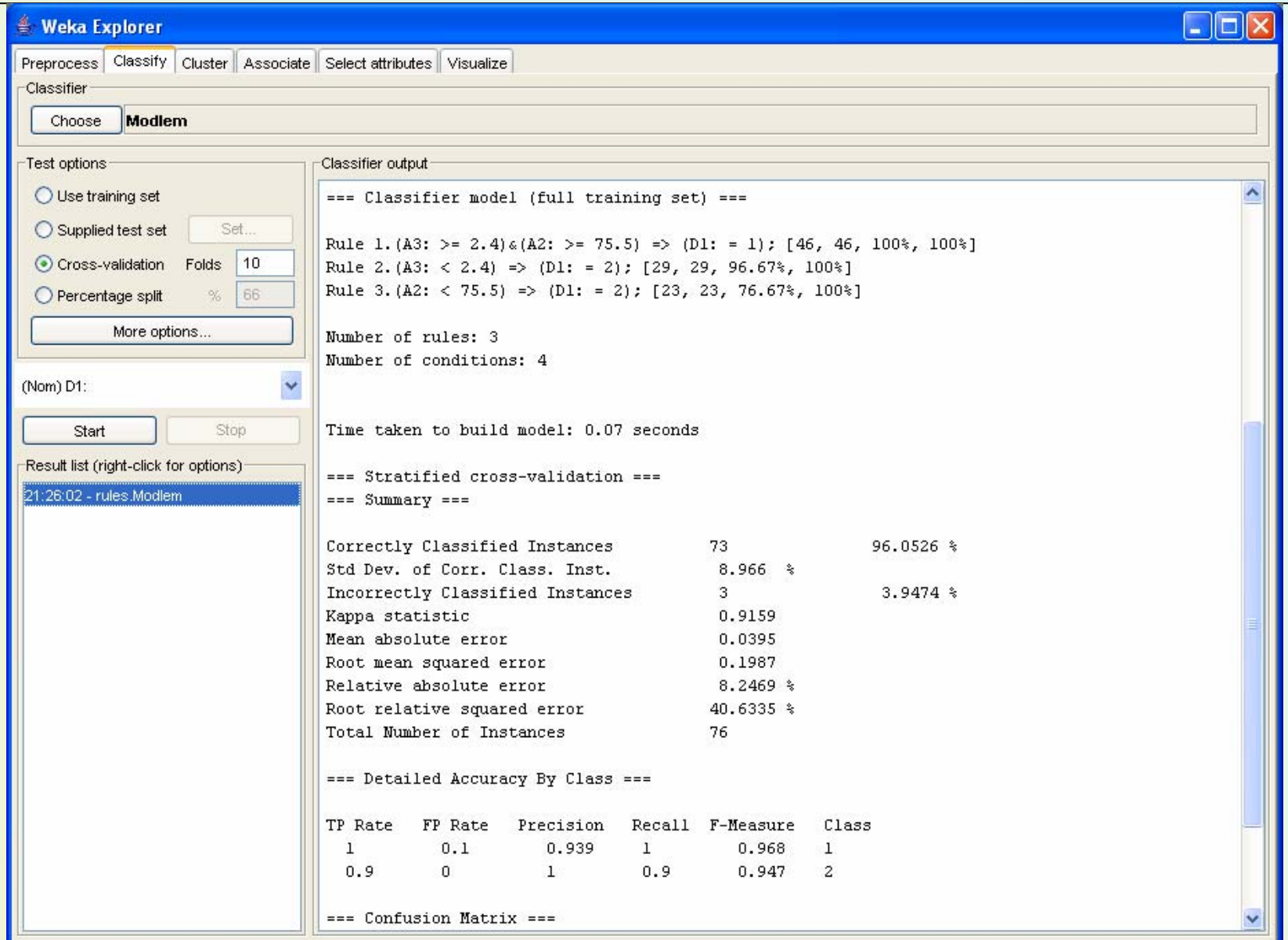
Drzewo J4.8 (WEKA)



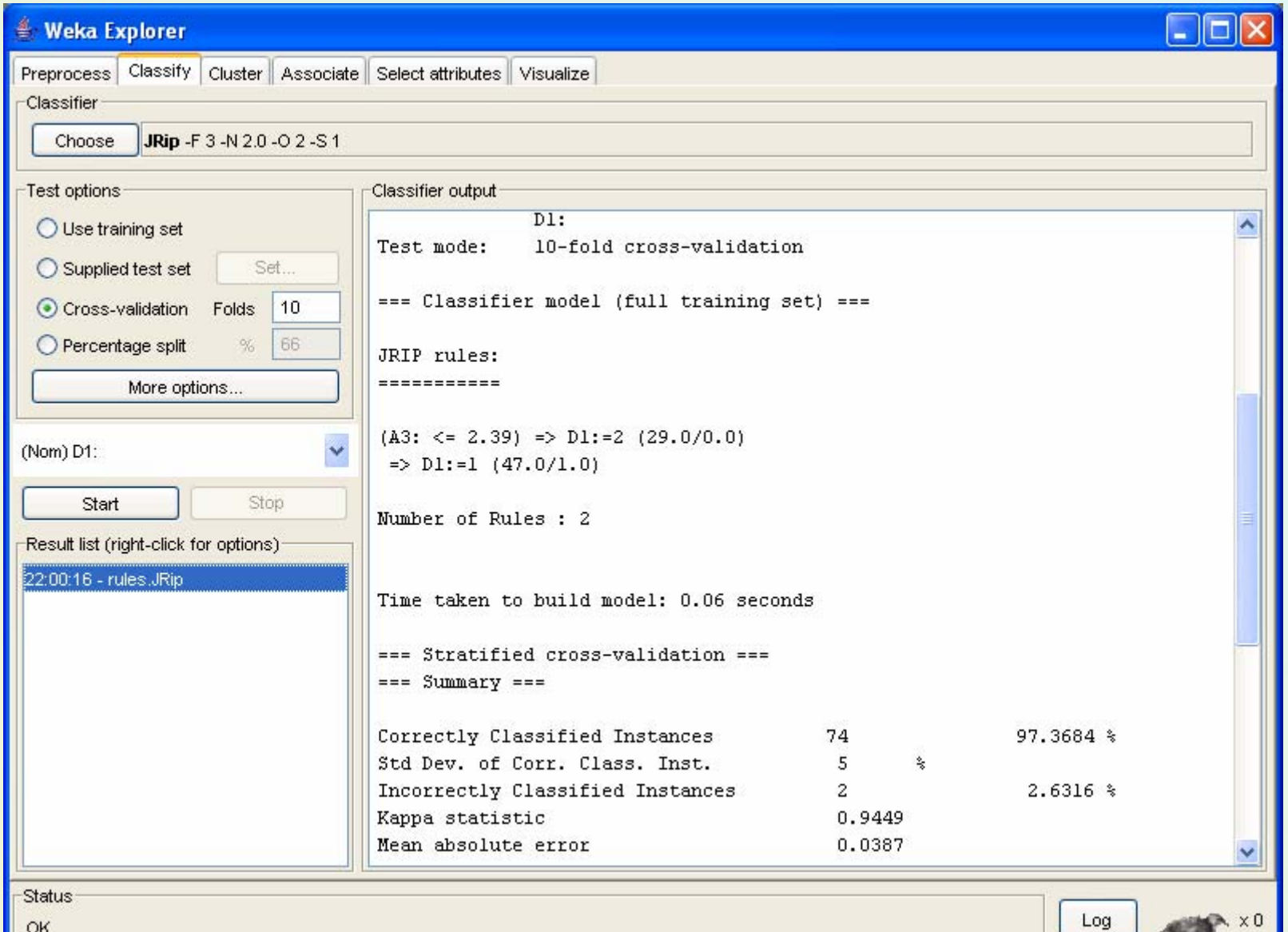
C4.5 z trochę innymi parametrami



Algorytm MODLEM



Ripper – alternatywne podejście do indukcji reguł

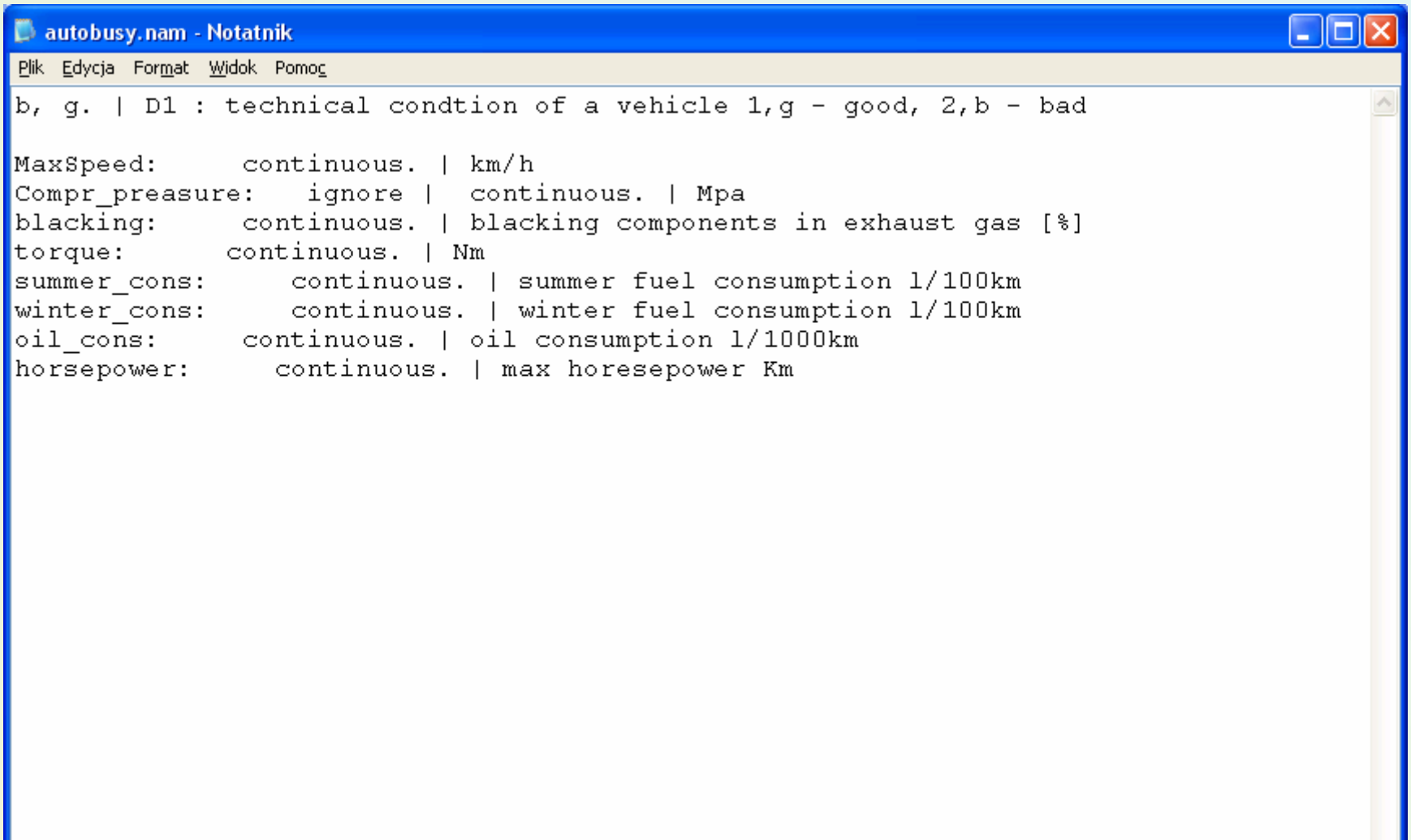


Podsumowanie klasyfikatorów

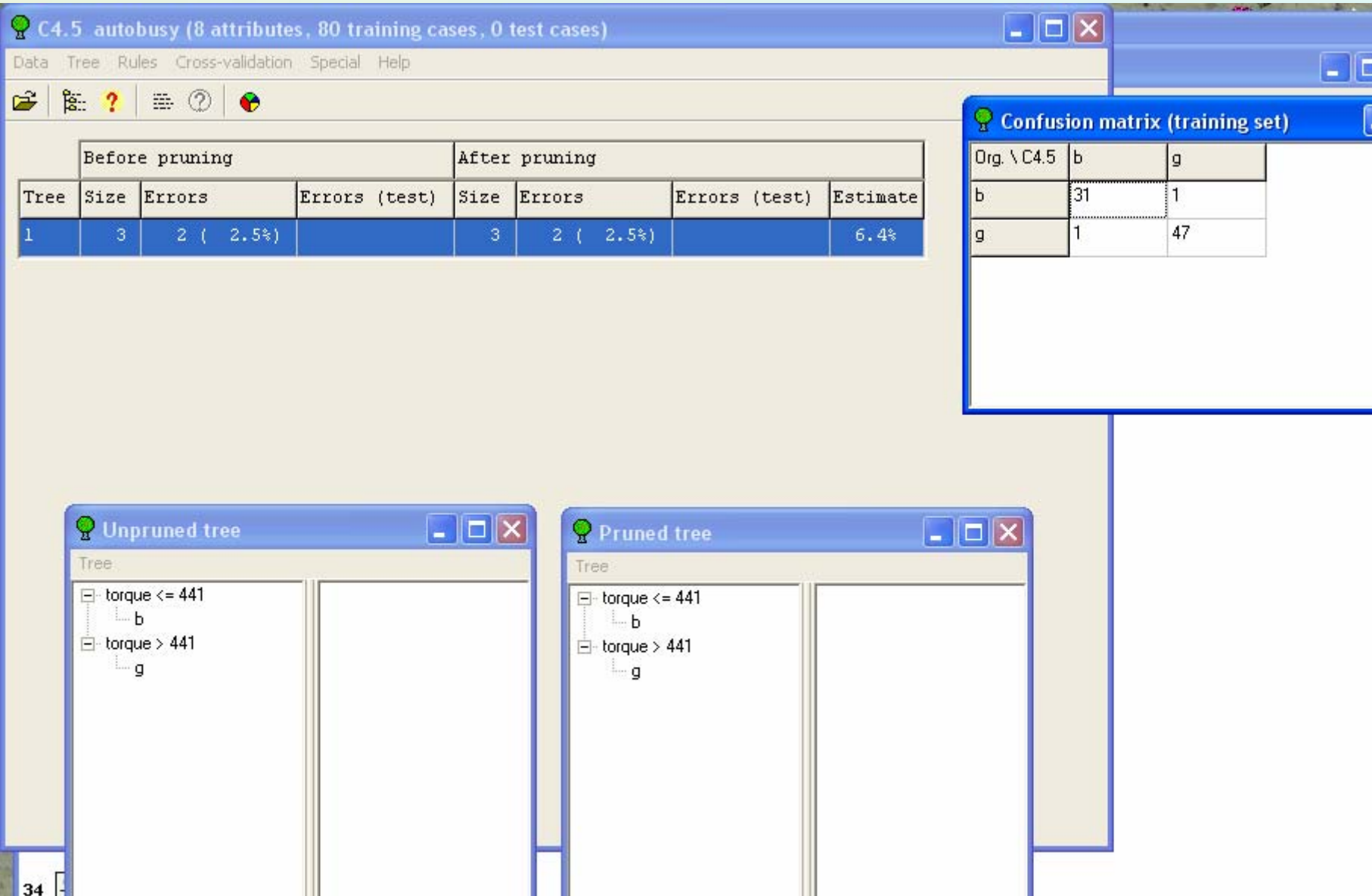
- Dość efektywny klasyfikator
  - Trafności około 97-98%
  - Rozpoznawanie trudniejszej klasy 90%
- Klasyfikatory symboliczne porównywalne do niesymbolicznych:
  - MLP BP – 97,37%
  - SVM – 97,37% (klasa n 0,93)
  - RBF – 93,27%
  - IBL (3) – 97,37%
  - J4.8 – 98,68%
- Można dokonać oceny wiedzy symbolicznej
- Podobną analizę warto przeprowadzić dla drugiej klasyfikacji.

Trudności pomiaru → alternatywne reprezentacje wiedzy klasyfikującej

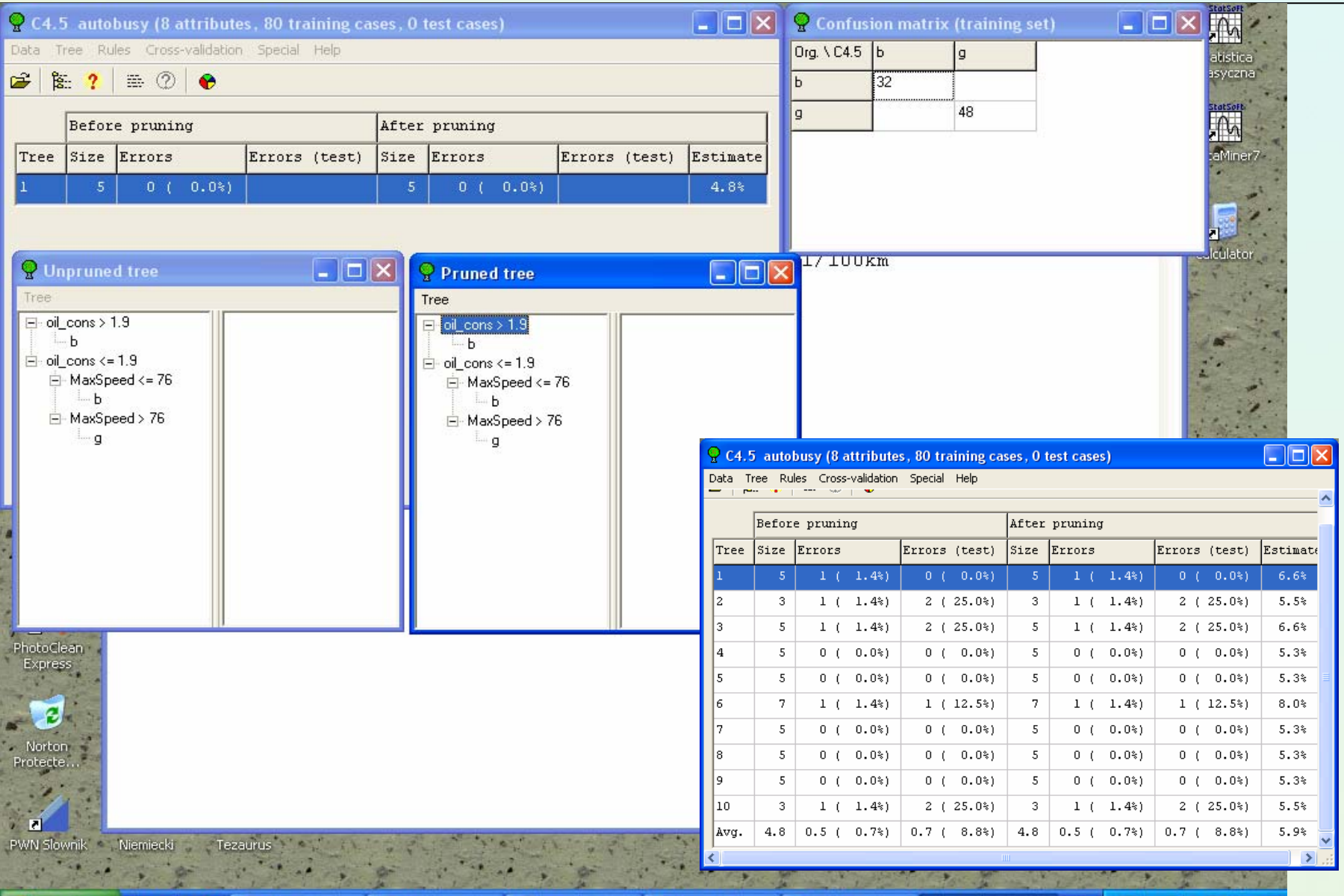
- Ciśnienie sprężania → najtrudniejsze do pomiaru



Pomijanie trudnych atrybutów (- ciśnienie sprężania)



Maskujmy dalsze atrybuty



Jeszcze inne możliwości

- Dalsze maskowania atrybutów w drzewie prowadzą do niższych trafności
- Jakie jeszcze metody oferują symboliczną reprezentacje wiedzy?
- Jak stworzyć profil/charakterystykę autobusu należącego do określonej klasy stanu technicznego?

Minimalny zbiór reguł klasyfikujących (MODLEM)

1. if (*s2*≥2.4 MPa) & (*s7*<2.1 *l*/1000km) then (technical state=good) [46]

2. if (*s2*<2.4 MPa) then (technical state=bad) [29]

3. if (*s7*≥2.1 *l*/1000km) then (technical state=bad) [24]

Oszacowana trafność klasyfikowania
('leaving one out' test) 98.7%.

Explore algorithm (Stefanowski, Vanderpooten)

- Inny cel poszukiwania reguł
  - to extract from data set inducing **all rules** that *satisfy* some *user's requirements* connected with *his interest* (regarding, e.g. the strength of the rule, level of confidence, length, sometimes also emphasis on the syntax of rules).
- Special technique of exploration the space of possible rules:
  - Progressively generation rules of increasing size using in the most efficient way some 'good' pruning and stopping condition that reject unnecessary candidates for rules.
- Similar to adaptations of Apriori principle for looking frequent itemsets [AIS94]; Brute [Etzioni]
- Więcej: J.Stefanowski, D.Vanderpooten: Induction of decision rules in classification and discovery-oriented perspectives, *International Journal of Intelligent Systems*, vol. 16 no. 1, 2001, 13-28.
- Lub monografia J.Stefanowski: Algorytmy indukcji regul decyzyjnych w odkrywaniu wiedzy => patrz http://www.cs.put.poznan.pl/jstefanowski/jspspdf.html

Poszukiwanie zbioru reguł silnych (EXPLORE)

Próg satysfakcji (51%):

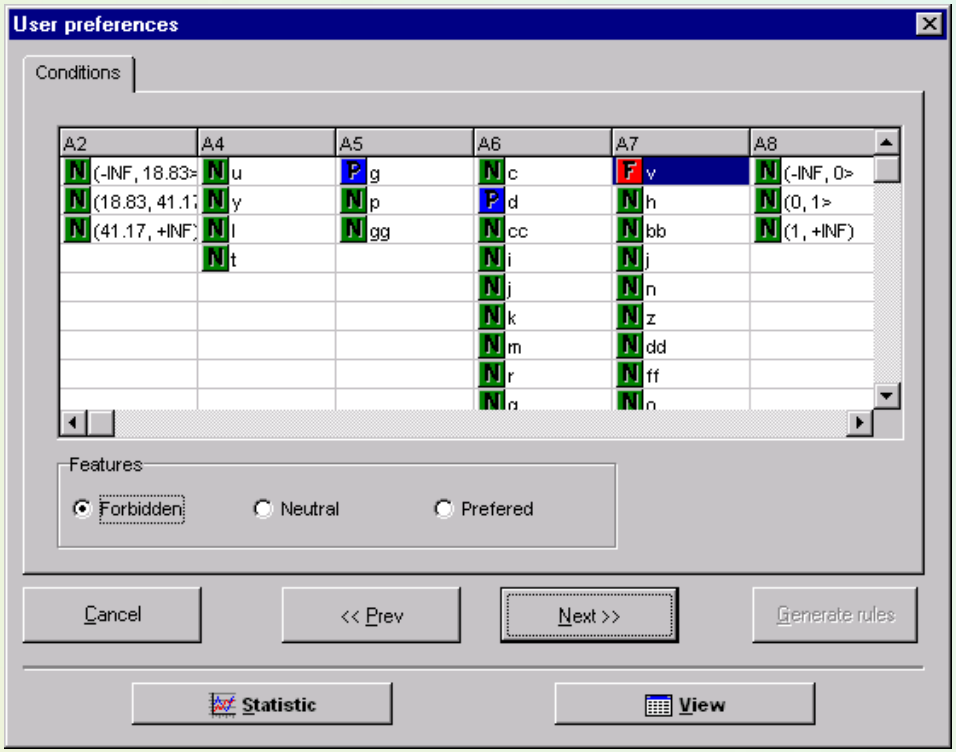
1. if (*s1*>85 km/h) then (technical state=good) [34]
2. if (*s8*>134 kM) then (technical state=good) [26]
3. if (*s2*≥2.4 MPa) & (*s3*<61 %) then (technical state=good) [44]
4. if (*s2*≥2.4 MPa) & (*s4*>444 Nm) then (technical state=good) [44]
5. if (*s2*≥2.4 MPa) & (s7<2.1 *l*/1000km) then (technical state=good) [46]
6. if (*s3*<61 %) & (*s4*>444 Nm) then (technical state=good) [42]
7. if (*s1*≤77 km/h) then (technical state=bad) [25]
8. if (*s2*<2.4 MPa) then (technical state=bad) [29]
9. if (*s7*≥2.1 *l*/1000km) then (technical state=bad) [24]
10. if (*s3*≥61 %) & (*s4*≤444 Nm) then (technical state=bad) [28]
11. if (*s3*³61 %) & (*s8*<120 kM) then (technical state=bad) [27]

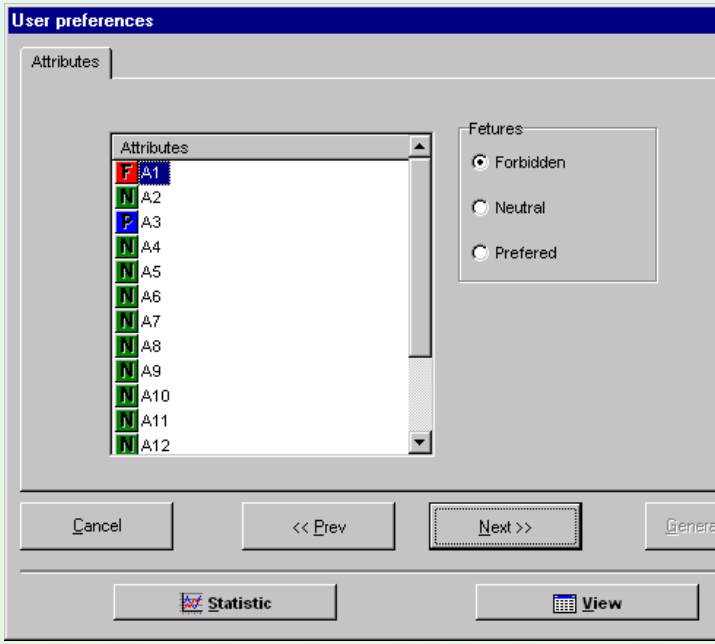
Możliwość dodatkowego ograniczania

- Interactive Explore
- Sterowanie procedurą poszukiwania reguł decyzyjnych za pomocą następujących parametrów :
  - ustalenie atrybutu decyzyjnego i jego klas decyzyjnych,
  - zbiór preferowanych i zakazanych atrybutów warunkowych,
  - zbiór preferowanych i zakazanych warunków elementarnych,
  - preferowany warunek złożony (część przesłanki),
  - zbiór zakazanych warunków złożonych,
  - maksymalna wielkość zbioru znalezionych reguł,
  - minimalne wsparcie reguły,
  - maksymalna długość reguły,
  - minimalny stopień dyskryminacji (dokładność) reguły.

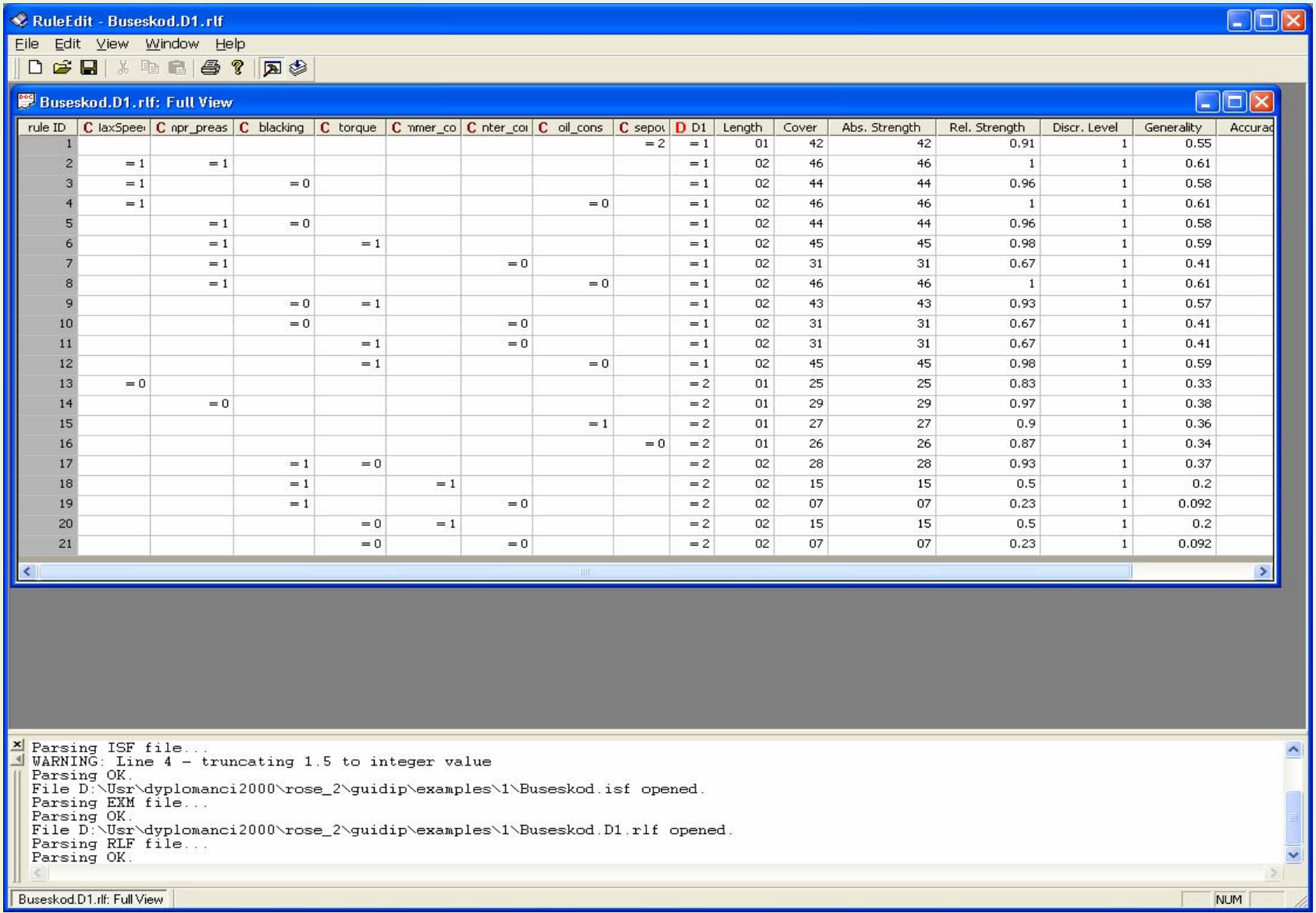
Sterowanie wyborem

- Interactive Explores





Przeglądanie zbioru reguł



rule ID	C laxSpee	C npr_preas	C blacking	C torque	C mmer_co	C nter_coi	C oil_cons	C sepo	D D1	Length	Cover	Abs. Strength	Rel. Strength	Discr. Level	Generality	Accurac
1								= 2	= 1	01	42	42	0.91	1	0.55	
2	= 1	= 1							= 1	02	46	46	1	1	0.61	
3	= 1		= 0						= 1	02	44	44	0.96	1	0.58	
4	= 1						= 0		= 1	02	46	46	1	1	0.61	
5		= 1	= 0						= 1	02	44	44	0.96	1	0.58	
6		= 1		= 1					= 1	02	45	45	0.98	1	0.59	
7		= 1				= 0			= 1	02	31	31	0.67	1	0.41	
8		= 1					= 0		= 1	02	46	46	1	1	0.61	
9			= 0	= 1					= 1	02	43	43	0.93	1	0.57	
10			= 0			= 0			= 1	02	31	31	0.67	1	0.41	
11				= 1		= 0			= 1	02	31	31	0.67	1	0.41	
12				= 1			= 0		= 1	02	45	45	0.98	1	0.59	
13	= 0								= 2	01	25	25	0.83	1	0.33	
14		= 0							= 2	01	29	29	0.97	1	0.38	
15							= 1		= 2	01	27	27	0.9	1	0.36	
16								= 0	= 2	01	26	26	0.87	1	0.34	
17			= 1	= 0					= 2	02	28	28	0.93	1	0.37	
18			= 1		= 1				= 2	02	15	15	0.5	1	0.2	
19			= 1			= 0			= 2	02	07	07	0.23	1	0.092	
20				= 0	= 1				= 2	02	15	15	0.5	1	0.2	
21				= 0		= 0			= 2	02	07	07	0.23	1	0.092	

Wymagania do sprawozdania

- Krótki opis problemu (raczej Twoje zrozumienie + cele postawione przez klienta)
- Informacje o danych
- Scenariusz metodyczny
- Wstępna ocena danych i ew. czyszczenie
- Analiza statystyk opisowych / korelacji itp./
- Ocena ważności atrybutów
  - Ranking
  - Ew. selekcja / redukcja
- Podsumowanie wiedzy klasyfikacyjnej
  - ***Nie tylko*** trafności
  - Próba interpretacji reprezentacji wiedzy
- Wnioski dla klienta

Źródło danych do problemu

- Dane zebrane przez dr hab. inż. J.Żak Wydziału maszyn Roboczych i Pojazdów Politechniki Poznańskiej
- J. Żak, J. Stefanowski. Determining maintenance activities of motor vehicles using rough sets aproach. In *Precedings of Euromaintenance'94 Conference*. Amsterdam 1994, 39 – 42.

O jakie autobusy chodziło

- Autosan H9-21(prototyp 1969, produkowany lata 73-84) w Sanockiej Fabryce Autobusów .

- Podstawowy model autobusu międzymiastowego PKS w poprzednim okresie