

Raport - Zaawansowane metody klasyfikacji oraz analiza skupień – algorytmy grupujące i hierarchiczne

Filip Michewicz 282239
Wiktor Niedźwiedzki 258882

10 czerwca 2025 Anno Domini

Spis treści

1	Zaawansowane metody klasyfikacji	3
1.1	Rodziny klasyfikatorów/uczenie zespołowe	3
1.1.1	Drzewa klasyfikacyjne	3
1.2	Metoda wektorów nośnych (SVM)	3
1.2.1	Jądro liniowe	3
1.2.2	Jądro wielomianowe	4
1.2.3	Jądro radialne	6
1.2.4	Jądro sigmoidalne	7
2	Analiza skupień – algorytmy grupujące i hierarchiczne	9
2.1	Zbiór danych Glass	9
2.2	Wyniki grupowania	10
2.2.1	k-średnie	11
2.2.2	Partitioning Around Medoids (PAM)	12
2.2.3	Agglomerative Nesting (AGNES)	14
2.2.4	Divisive clustering (DIANA)	17

Spis wykresów

1	Wykres pudełkowy, zmienne bez standaryzacji	9
2	Wykres pudełkowe, po standaryzacji	10
3	Wizualizacja danych, PCA	10
4	PCA, kolory - rzeczywiste, kształt - wyniki	11
5	Wykres RI od Na, aby pokazać gdzie są wyznaczone centra skupień	11
6	coś	12
7	coś, z medoidami	13
8	AGNES: single linkage	14
9	AGNES: complete linkage	15
10	AGNES: average linkage	16
11	AGNES: average linkage	17
12	AGNES: average linkage	18

Spis tabel

1	Średnia poprawa dokładności klasyfikacji za pomocą drzewa klasyfikacyjnego, z podziałem na algorytmy uczenia zespołowego oraz liczbę replikacji	3
---	---	---

2	Jądro liniowe - bez skalowania	3
3	Jądro liniowe - ze skalowaniem	4
4	Jądro wielomianowe - wielokrotny podział, bez skalowania	4
5	Jądro wielomianowe - wielokrotny podział, ze skalowaniem	4
6	Jądro wielomianowe - cross-validation, bez skalowania	4
7	Jądro wielomianowe - cross-validation, ze skalowaniem	5
8	Jądro wielomianowe - bootstrap, bez skalowania	5
9	Jądro wielomianowe - bootstrap, ze skalowaniem	5
10	Badanie wpływu stopnia wielomianu na dokładność - wielokrotny podział, najbardziej dokładna kombinacja gammy i kary dla opcji default (stopień 3	6
11	Jądro radialne - wielokrotny podział, bez skalowania	6
12	Jądro radialne - wielokrotny podział, ze skalowaniem	6
13	Jądro radialne - cross-validation, bez skalowania	6
14	Jądro radialne - cross-validation, ze skalowaniem	7
15	Jądro radialne - bootstrap, bez skalowania	7
16	Jądro radialne - bootstrap, ze skalowaniem	7
17	Jądro sigmoidalne - wielokrotny podział, bez skalowania	7
18	Jądro sigmoidalne - wielokrotny podział, ze skalowaniem	8
19	Jądro sigmoidalne - cross-validation, bez skalowania	8
20	Jądro sigmoidalne - cross-validation, ze skalowaniem	8
21	Jądro sigmoidalne - bootstrap, bez skalowania	8
22	Jądro sigmoidalne - bootstrap, ze skalowaniem	9
23	Macierz błędów; metoda k-średnich	11
24	Macierz błędów; metoda k-średnich	13
25	Macierz błędów; agnes, najbliższy sąsiad	14
26	Macierz błędów; agnes, najdalszy sąsiad	15
27	Macierz błędów; agnes, średnia odległość	16
28	Macierz błędów; agnes, średnia odległość	17
29	Macierz błędów; agnes, średnia odległość	18

1 Zaawansowane metody klasyfikacji

W pierwszej części zadania zastosujemy algorytmy *ensemble learning* (bagging, boosting i random forest) w celu poprawy dokładności cech klasyfikacyjnych. W drugiej natomiast poznamy i ocenimy nową metodę klasyfikacji - metodę wektorów nośnych (SVM).

Zadanie zostanie wykonane na zbiorze danych *wine*, którego szczegółowy opis znajduje się w poprzednim raporcie.

1.1 Rodziny klasyfikatorów/uczenie zespołowe

Wyróżniamy trzy algorytmy uczenia zespołowego (ang. ensemble learning):

- **Bagging** - generujemy B-bootstrapowych replikacji zbioru uczącego, na podstawie których tworzymy B klasyfikatorów. Następnie łączymy je w klasyfikator zagregowany, który przydziela dane cechy do klas za pomocą reguły "głosowania większości" (w przypadku remisu wybiera losowo). Każdy klasyfikator powstaje niezależnie (w sensie takim, że wyniki poprzednich nie mają wpływu na generowanie nowych).
- **boosting** - podobnie jak w bagging, tworzymy klasyfikator zagregowany złożony z wielu pojedynczych klasyfikatorów. Jednak różnica jest taka, że klasyfikatory powstają sekwencyjnie. Na początku każda cecha w zbiorze ma przypisaną taką samą wagę. Z każdą kolejną iteracją natomiast waga zwiększa się dla uprzednio źle sklasyfikowanych przypadków.
- **random forest** (dla drzew klasyfikacyjnych) - metoda podobna do bagging z tą różnicą, że klasyfikatory powstają na podstawie różnych m-elementowych podzbiorach cech (m mniejsze bądź równe wszystkim cechom).

1.1.1 Drzewa klasyfikacyjne

MOŻE BYĆ NIEPOOPRAWNIE W CHUJ

Tabela 1: Średnia poprawa dokładności klasyfikacji za pomocą drzewa klasyfikacyjnego, z podziałem na algorytmy uczenia zespołowego oraz liczbę replikacji

	1	5	10	20	30	40	50	100
Bagging	44.95	35.31	59.41	59.63	63.57	71.13	59.02	62.82
Random Forest	86.64	89.39	79.93	86.27	88.43	88.67	86.37	85.99
Boosting	77.54	77.77	70.30	69.48	75.88	73.91	68.71	71.24
Średnia	69.71	67.49	69.88	71.79	75.96	77.90	71.37	73.35

1.2 Metoda wektorów nośnych (SVM)

W tej części przeprowadzona będzie klasyfikacja na podstawie metody wektorów nośnych, z podziałem na różne funkcje jądrowe.

COŚ O TYM CO TO WOGÓLE JEST

1.2.1 Jądro liniowe

Tabela 2: Jądro liniowe - bez skalowania

	0.001	0.01	0.1	1	10	100	1000
Wielokrotny podział	42.67	97.17	98.17	96.83	95.33	95.83	95.00
Cross-validation	34.52	44.70	91.51	91.14	91.14	91.14	91.14
Bootstrap	37.94	97.25	97.17	96.55	95.75	96.44	96.30

Tabela 2: Jądro liniowe - bez skalowania (kontynuacja)

	0.001	0.01	0.1	1	10	100	1000
Średnio	38.37	79.71	95.62	94.84	94.07	94.47	94.15

Tabela 3: Jądro liniowe - ze skalowaniem

	0.001	0.01	0.1	1	10	100	1000
Wielokrotny podział	41.17	97.00	98.00	95.67	95.33	96.17	96.50
Cross-validation	33.96	49.82	92.38	92.07	92.07	92.07	92.07
Bootstrap	32.91	97.47	96.57	97.14	97.45	96.83	97.54
Średnio	36.01	81.43	95.65	94.96	94.95	95.02	95.37

1.2.2 Jądro wielomianowe

Tabela 4: Jądro wielomianowe - wielokrotny podział, bez skalowania

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	39.17	96.67	95.33	97.00	95.33	94.67	96.33	95.17	95.67	95.67
0.01	38.33	94.67	96.33	96.67	96.00	96.17	95.00	94.67	95.33	95.17
0.1	42.67	95.50	97.00	95.67	96.67	93.17	95.67	94.67	94.67	95.83
1	37.67	94.83	95.83	96.17	93.00	94.50	96.00	96.17	95.00	95.17
10	37.83	94.17	94.83	94.33	96.00	96.00	95.50	97.17	94.50	93.83
100	84.83	95.83	95.50	95.17	94.50	95.17	94.67	96.00	95.67	96.00
1000	94.83	96.33	95.83	95.67	95.33	95.67	95.00	95.00	96.67	95.00

Tabela 5: Jądro wielomianowe - wielokrotny podział, ze skalowaniem

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	37.50	95.83	94.83	95.33	95.50	95.50	95.50	93.50	95.50	94.50
0.01	38.83	95.33	95.50	96.17	96.00	95.00	96.50	94.17	96.17	95.67
0.1	38.83	96.17	95.00	94.83	93.33	93.83	95.50	95.83	95.67	96.17
1	39.17	95.67	96.50	96.00	94.83	95.33	95.17	95.33	96.83	95.00
10	38.67	94.67	95.83	95.17	96.83	95.33	93.50	92.83	95.00	95.67
100	84.17	95.50	95.83	95.00	95.00	96.00	95.50	94.83	95.67	96.33
1000	95.50	94.00	96.50	94.83	95.33	95.83	95.50	96.33	94.33	96.00

Tabela 6: Jądro wielomianowe - cross-validation, bez skalowania

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	39.90	95.52	96.08	96.08	96.08	96.08	96.08	96.08	96.08	96.08
0.01	39.90	96.08	96.08	96.08	96.08	96.08	96.08	96.08	96.08	96.08
0.1	39.90	96.08	96.08	96.08	96.08	96.08	96.08	96.08	96.08	96.08
1	39.90	96.08	96.08	96.08	96.08	96.08	96.08	96.08	96.08	96.08
10	42.71	96.08	96.08	96.08	96.08	96.08	96.08	96.08	96.08	96.08

Tabela 6: Jądro wielomianowe - cross-validation, bez skalowania
(kontynuacja)

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
100	87.06	96.08	96.08	96.08	96.08	96.08	96.08	96.08	96.08	96.08
1000	96.08	96.08	96.08	96.08	96.08	96.08	96.08	96.08	96.08	96.08

Tabela 7: Jądro wielomianowe - cross-validation, ze skalowaniem

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	39.90	96.05	96.6	96.6	96.6	96.6	96.6	96.6	96.6	96.6
0.01	39.90	96.60	96.6	96.6	96.6	96.6	96.6	96.6	96.6	96.6
0.1	39.90	96.60	96.6	96.6	96.6	96.6	96.6	96.6	96.6	96.6
1	39.90	96.60	96.6	96.6	96.6	96.6	96.6	96.6	96.6	96.6
10	44.38	96.60	96.6	96.6	96.6	96.6	96.6	96.6	96.6	96.6
100	88.89	96.60	96.6	96.6	96.6	96.6	96.6	96.6	96.6	96.6
1000	96.63	96.60	96.6	96.6	96.6	96.6	96.6	96.6	96.6	96.6

Tabela 8: Jądro wielomianowe - bootstrap, bez skalowania

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	37.63	95.61	94.69	95.49	93.61	95.40	94.01	94.43	95.33	94.04
0.01	37.77	94.78	93.95	94.17	92.99	93.43	92.82	94.35	94.23	94.05
0.1	39.00	94.90	95.12	94.76	95.12	94.16	93.92	95.54	95.49	94.16
1	36.64	94.01	95.79	94.50	94.60	94.92	95.10	94.89	93.51	95.05
10	38.55	95.83	93.97	93.79	94.68	94.98	95.15	94.66	94.64	94.07
100	82.92	94.43	95.64	95.25	94.48	94.01	93.72	93.59	94.44	93.91
1000	94.50	94.94	95.30	93.49	93.82	94.18	94.21	93.48	95.35	95.39

Tabela 9: Jądro wielomianowe - bootstrap, ze skalowaniem

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	32.93	93.03	94.89	93.42	94.43	93.06	93.88	93.67	93.39	94.17
0.01	37.47	92.84	93.37	93.51	94.33	94.86	94.90	95.29	95.39	93.17
0.1	36.74	94.02	95.41	94.75	95.07	95.50	96.03	93.86	94.69	95.38
1	37.28	94.43	93.52	94.53	94.85	93.56	94.97	94.77	95.40	93.45
10	39.87	95.84	94.56	95.96	96.23	93.18	93.63	94.53	94.60	94.04
100	85.29	93.64	94.50	93.95	95.09	95.15	94.44	93.69	93.01	95.64
1000	94.36	95.23	95.16	94.54	94.11	94.96	94.46	94.72	94.26	94.51

Najlepsza gamma: **8.89**, najlepsza kara: **1**. Robimy dla danych po standaryzacji, bo tak i chuj.

Badamy tylko na podstawie wielokrotnego podziału, bo tak i chuj również.

Tabela 10: Badanie wpływu stopnia wielomianu na dokładność - wielokrotny podział, najbardziej dokładna kombinacja gammy i kary dla opcji default (stopień 3)

	2	3	4	5	6	7
Dokładność	87.33	94.33	83.33	88.33	80.17	82

1.2.3 Jądro radialne

Tabela 11: Jądro radialne - wielokrotny podział, bez skalowania

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87
0.01	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87
0.1	79.80	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87
1	97.71	56.18	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87
10	97.78	61.80	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87
100	96.63	61.80	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87
1000	96.63	61.80	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87

Tabela 12: Jądro radialne - wielokrotny podział, ze skalowaniem

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	40.07	40.07	40.07	40.07	40.07	40.07	40.07	40.07	40.07	40.07
0.01	40.07	40.07	40.07	40.07	40.07	40.07	40.07	40.07	40.07	40.07
0.1	81.01	40.07	40.07	40.07	40.07	40.07	40.07	40.07	40.07	40.07
1	98.86	51.86	40.07	40.07	40.07	40.07	40.07	40.07	40.07	40.07
10	98.30	59.71	40.65	40.07	40.07	40.07	40.07	40.07	40.07	40.07
100	97.12	59.71	40.65	40.07	40.07	40.07	40.07	40.07	40.07	40.07
1000	97.12	59.71	40.65	40.07	40.07	40.07	40.07	40.07	40.07	40.07

Tabela 13: Jądro radialne - cross-validation, bez skalowania

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	39.90	39.90	39.9	39.9	39.9	39.9	39.9	39.9	39.9	39.9
0.01	39.90	39.90	39.9	39.9	39.9	39.9	39.9	39.9	39.9	39.9
0.1	79.77	39.90	39.9	39.9	39.9	39.9	39.9	39.9	39.9	39.9
1	98.33	56.18	39.9	39.9	39.9	39.9	39.9	39.9	39.9	39.9
10	96.63	60.16	39.9	39.9	39.9	39.9	39.9	39.9	39.9	39.9
100	96.63	60.16	39.9	39.9	39.9	39.9	39.9	39.9	39.9	39.9
1000	96.63	60.16	39.9	39.9	39.9	39.9	39.9	39.9	39.9	39.9

Tabela 14: Jądro radialne - cross-validation, ze skalowaniem

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	39.77	39.77	39.77	39.77	39.77	39.77	39.77	39.77	39.77	39.77
0.01	39.77	39.77	39.77	39.77	39.77	39.77	39.77	39.77	39.77	39.77
0.1	80.95	39.77	39.77	39.77	39.77	39.77	39.77	39.77	39.77	39.77
1	98.33	55.10	39.77	39.77	39.77	39.77	39.77	39.77	39.77	39.77
10	97.75	60.72	39.77	39.77	39.77	39.77	39.77	39.77	39.77	39.77
100	97.75	60.72	39.77	39.77	39.77	39.77	39.77	39.77	39.77	39.77
1000	97.75	60.72	39.77	39.77	39.77	39.77	39.77	39.77	39.77	39.77

Tabela 15: Jądro radialne - bootstrap, bez skalowania

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	37.73	35.69	34.68	34.80	38.40	34.22	37.85	36.98	35.55	39.04
0.01	38.59	39.10	34.06	38.30	34.93	37.81	37.18	37.27	38.02	35.94
0.1	61.20	35.69	35.41	34.33	37.65	36.01	37.98	35.68	37.10	39.94
1	97.14	38.01	35.09	33.66	36.76	37.01	37.63	38.23	36.88	34.10
10	96.81	54.71	35.43	37.94	36.70	39.12	38.27	35.94	37.89	36.40
100	95.88	53.38	40.77	38.90	39.11	39.25	36.38	38.21	36.07	37.64
1000	96.75	57.47	38.47	37.66	36.73	38.12	37.86	36.22	36.62	35.94

Tabela 16: Jądro radialne - bootstrap, ze skalowaniem

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	35.71	35.17	35.94	40.62	37.69	38.48	34.85	37.71	36.01	37.45
0.01	35.11	38.57	36.08	36.79	36.05	36.85	40.82	36.31	35.66	40.74
0.1	54.26	36.97	38.83	40.31	35.45	35.87	37.35	35.91	36.03	36.35
1	97.73	47.93	40.94	38.98	33.98	38.71	37.83	38.32	36.76	36.59
10	97.38	50.60	38.80	38.45	36.58	34.43	37.81	37.34	39.02	35.50
100	97.28	53.52	34.98	37.82	38.62	41.22	39.88	40.57	36.78	38.66
1000	96.23	49.01	38.12	37.66	38.12	38.94	37.81	38.68	40.24	38.60

1.2.4 Jądro sigmoidalne

Tabela 17: Jądro sigmoidalne - wielokrotny podział, bez skalowania

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	38.83	36.00	37.17	39.83	39.17	44.00	39.00	39.33	41.17	41.17
0.01	40.17	66.00	57.17	60.33	62.67	61.83	58.50	69.83	63.67	64.50
0.1	41.67	89.33	85.83	88.83	86.17	86.17	86.00	87.83	88.33	87.83
1	97.17	83.83	78.17	78.17	81.00	82.83	84.33	80.83	82.67	80.83
10	97.33	80.17	79.33	79.00	80.00	78.83	80.83	83.33	81.33	77.33
100	96.50	82.67	81.67	79.50	81.17	80.67	78.50	83.33	81.00	79.83
1000	96.33	84.17	82.00	78.00	82.17	81.83	80.67	80.17	81.67	81.50

Tabela 18: Jądro sigmoidalne - wielokrotny podział, ze skalowaniem

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	44.00	40.50	40.67	35.33	41.33	41.83	38.17	41.83	36.83	37.50
0.01	39.50	60.83	62.83	62.00	59.83	61.33	60.50	64.67	58.33	61.50
0.1	37.17	87.17	86.33	88.50	87.67	88.00	88.00	86.83	86.17	86.83
1	98.33	84.00	82.17	81.67	83.17	80.00	83.00	79.00	79.33	85.33
10	97.50	81.00	81.33	84.00	82.67	83.33	83.33	80.67	81.83	81.33
100	96.33	78.33	81.50	83.83	78.50	81.50	81.67	81.83	79.67	80.83
1000	95.50	84.83	83.00	80.17	83.50	80.17	80.33	82.00	81.00	80.83

Tabela 19: Jądro sigmoidalne - cross-validation, bez skalowania

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87
0.01	39.87	80.85	80.85	80.85	80.29	80.29	80.29	80.29	80.29	80.85
0.1	42.19	88.30	85.49	84.90	84.35	84.90	84.35	86.01	84.90	85.46
1	98.33	84.38	81.01	78.17	76.54	77.71	77.71	77.65	77.12	76.01
10	97.78	85.42	82.06	78.66	76.99	77.09	77.12	78.17	78.24	78.82
100	96.05	87.09	82.06	78.69	75.88	78.20	79.35	78.14	77.12	77.68
1000	96.05	86.54	82.65	78.14	77.55	77.65	77.68	78.17	78.24	79.35

Tabela 20: Jądro sigmoidalne - cross-validation, ze skalowaniem

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	40.00	40.00	40.00	40.00	40.00	40.00	40.00	40.00	40.00	40.00
0.01	40.00	81.54	80.42	81.01	81.01	81.01	81.01	81.01	81.01	81.01
0.1	43.37	88.24	88.24	88.20	86.54	85.42	85.98	85.42	85.42	85.42
1	97.19	83.73	80.42	82.65	79.84	78.20	79.80	80.92	82.58	80.33
10	98.86	81.44	80.39	79.28	79.25	80.92	78.07	77.52	78.66	77.55
100	96.05	79.15	82.61	78.17	79.80	79.25	77.52	78.63	79.77	78.10
1000	96.05	79.15	85.95	78.17	77.03	80.36	74.74	75.85	77.55	78.66

Tabela 21: Jądro sigmoidalne - bootstrap, bez skalowania

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	34.96	38.32	37.39	36.28	39.07	39.62	41.16	34.41	42.59	39.32
0.01	36.96	58.27	58.19	63.44	69.59	62.37	61.47	65.49	64.63	57.54
0.1	39.00	85.33	87.85	85.88	84.57	82.79	84.24	84.66	85.02	85.03
1	97.29	80.48	77.61	80.83	82.06	79.64	79.63	81.62	80.33	76.06
10	96.64	77.96	80.95	77.27	80.54	79.08	77.24	77.86	76.77	82.53
100	97.20	79.02	81.91	81.65	78.91	77.21	82.12	80.42	81.59	77.19
1000	96.13	79.98	78.36	81.39	80.61	77.61	79.10	80.45	78.77	79.61

Tabela 22: Jądro sigmoidalne - bootstrap, ze skalowaniem

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	35.82	33.95	40.20	36.55	39.67	37.46	40.98	35.47	35.39	35.07
0.01	35.10	58.20	60.34	67.68	56.81	55.90	54.47	58.13	59.68	62.05
0.1	43.79	85.95	85.38	85.33	86.97	85.26	86.30	85.08	85.40	85.25
1	96.65	82.57	82.74	80.44	78.75	80.29	77.05	82.20	78.95	78.12
10	95.96	79.70	80.98	80.60	78.48	74.00	82.79	80.80	78.32	77.46
100	97.06	80.82	76.97	79.72	79.87	81.57	80.47	78.19	81.67	80.42
1000	97.18	81.36	78.01	79.71	78.95	78.20	79.88	78.65	79.91	78.24

elo

2 Analiza skupień – algorytmy grupujące i hierarchiczne

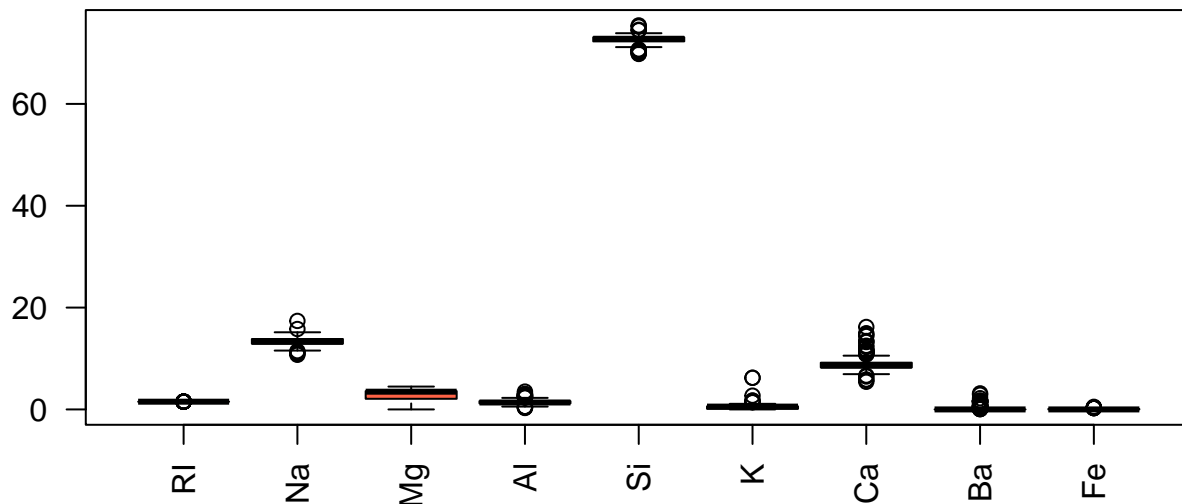
W pierwszej części zadania zastosujemy algorytmy *ensemble learning* (bagging, boosting i random forest) w celu poprawy dokładności cech klasyfikacyjnych. W drugiej natomiast poznamy i ocenimy nową metodę klasyfikacji - metodę wektorów nośnych (SVM).

Zadanie zostanie wykonane na zbiorze danych *wine*, którego szczegółowy opis znajduje się w poprzednim raporcie.

W tym zadaniu zastosujemy i porównamy ze sobą metody analizy skupień - k-średnich i PAM jako algorytmy grupujące, oraz AGNES - algorytm hierarchiczny.

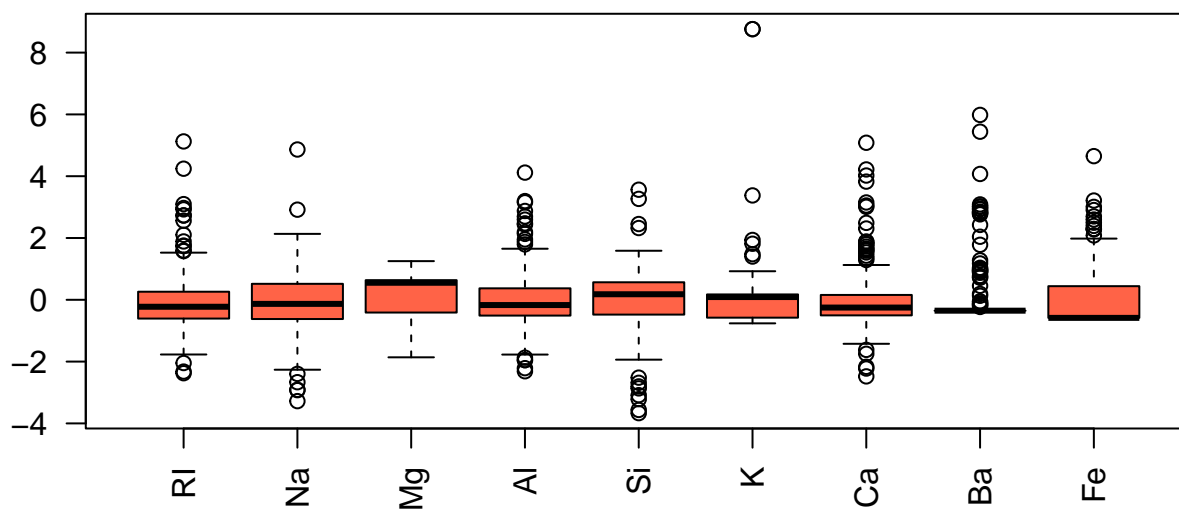
2.1 Zbiór danych Glass

Chujowy



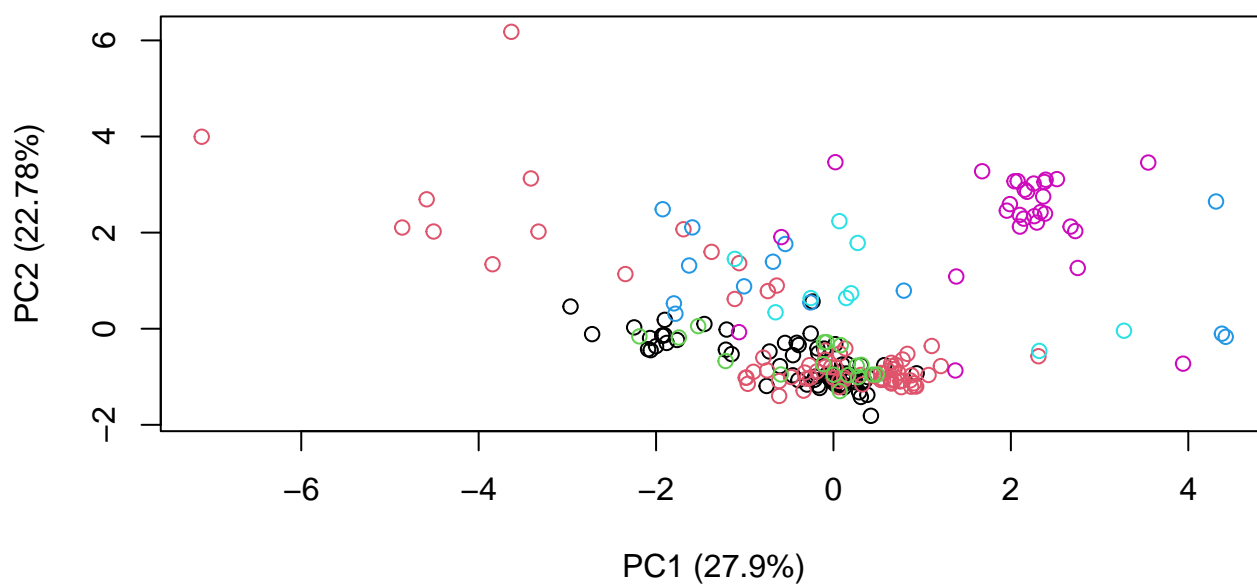
Wykres 1: Wykres pudełkowy, zmienne bez standaryzacji

KURWA STANDARYZACJA MACHEN



Wykres 2: Wykres pudełkowe, po standaryzacji

Teraz jest zajebicie



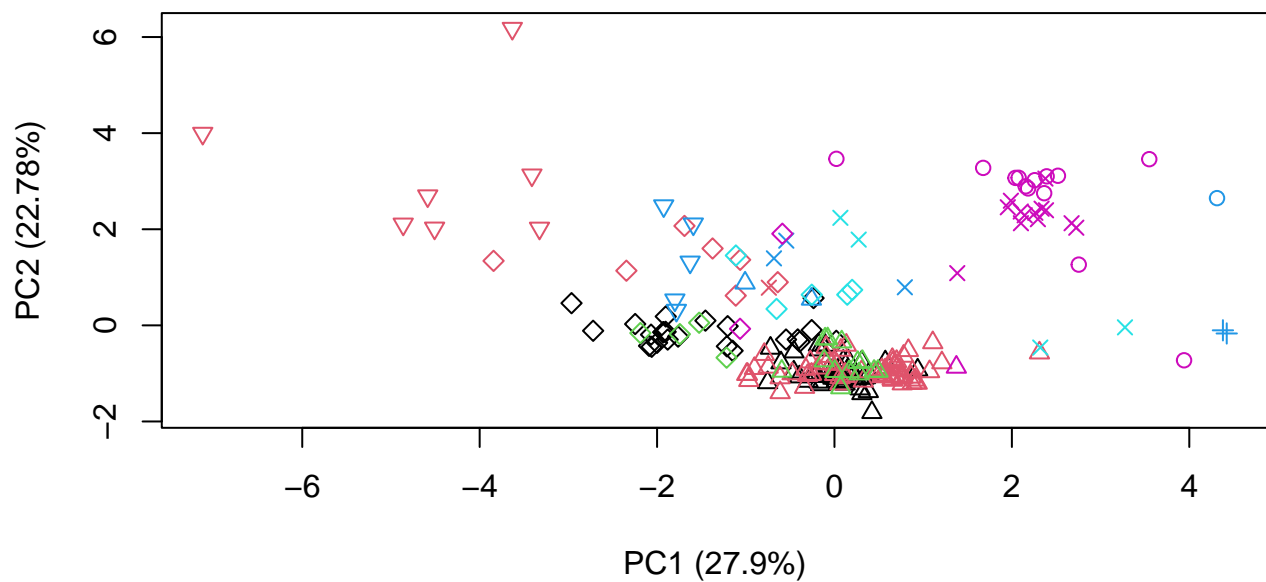
Wykres 3: Wizualizacja danych, PCA

Chuja widać, ciekawe kto wybrał ten zbiór?

2.2 Wyniki grupowania

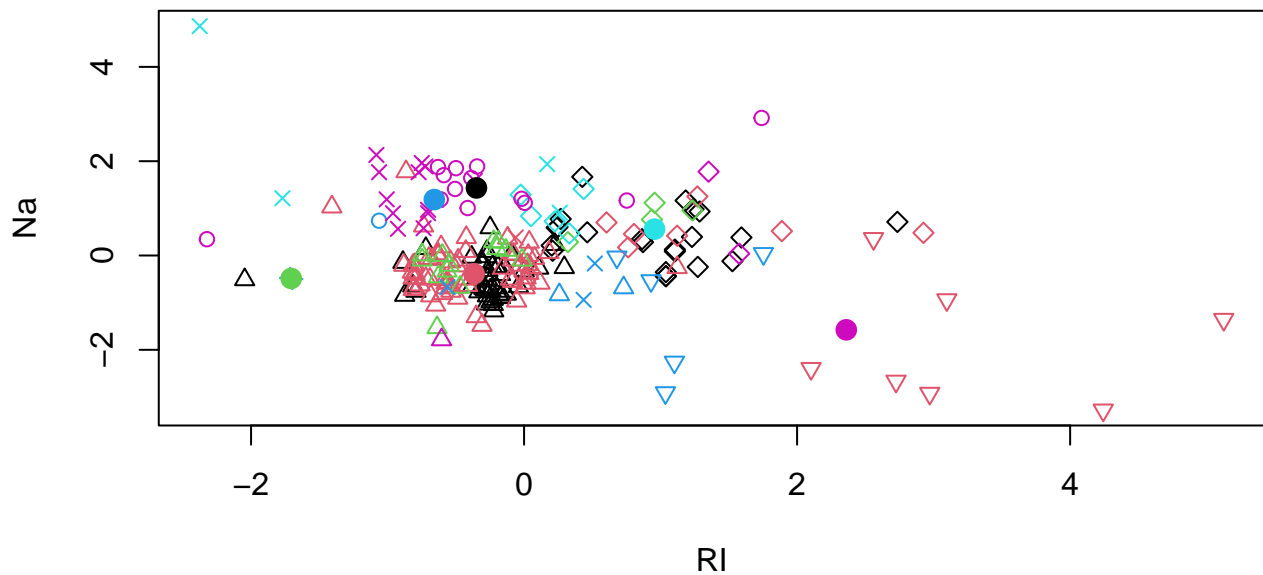
Przeprowadzamy dla rzeczywistej liczby etykiet, która wynosi **6**.

2.2.1 k-średnie



Wykres 4: PCA, kolory - rzeczywiste, kształt - wyniki

Gównianie mu poszło



Wykres 5: Wykres RI od Na, aby pokazać gdzie są wyznaczone centra skupień

Centra wywalone w kosmos, ale fajnie

Tabela 23: Macierz błędów; metoda k-średnich

	1	2	3	5	6	7
1	23	7	4	0	5	2

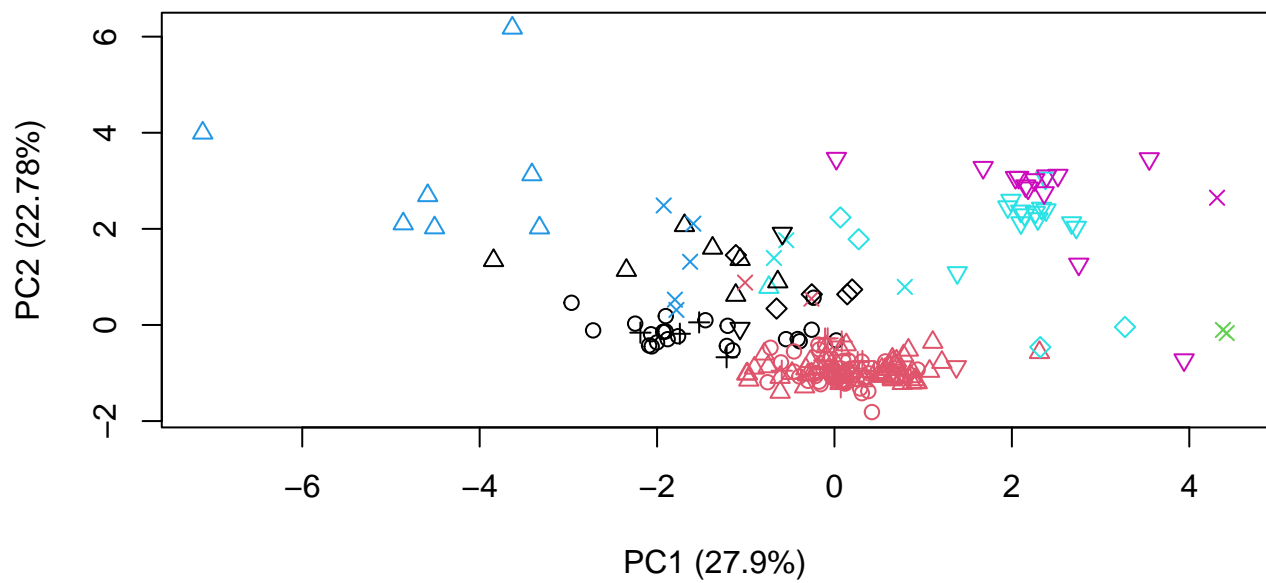
	1	2	3	5	6	7
2	47	61	13	2	0	1
3	0	0	0	2	0	0
5	0	7	0	5	0	0
6	0	1	0	3	4	13
7	0	0	0	1	0	13

Dokładność: 49.53%.

2.2.2 Partitioning Around Medoids (PAM)

```
## Error: nie znaleziono obiektu 'conf_mat'
```

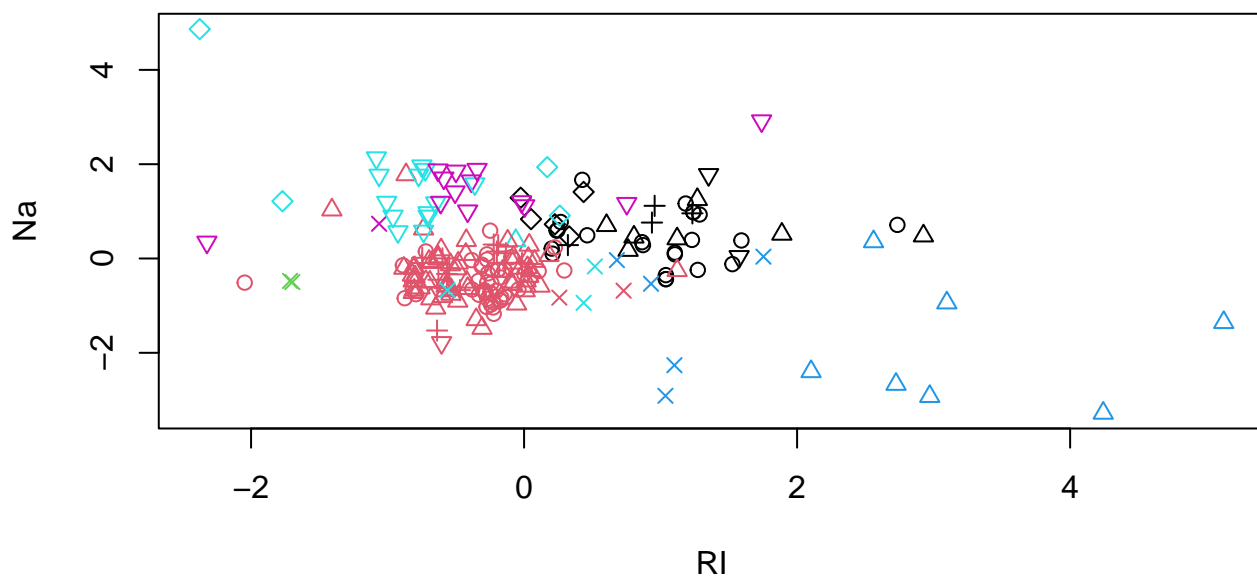
```
## Error: nie znaleziono obiektu 'glass.pam7'
```



Wykres 6: coś

Też słabo

```
## Error: nie znaleziono obiektu 'glass.pam7'
```



Wykres 7: coś, z medoidami

Error in medoids[, c("RI", "Na")]: obiekt typu 'closure' nie jest ustawialny
Meh

Tabela 24: Macierz błędów; metoda k-średnich

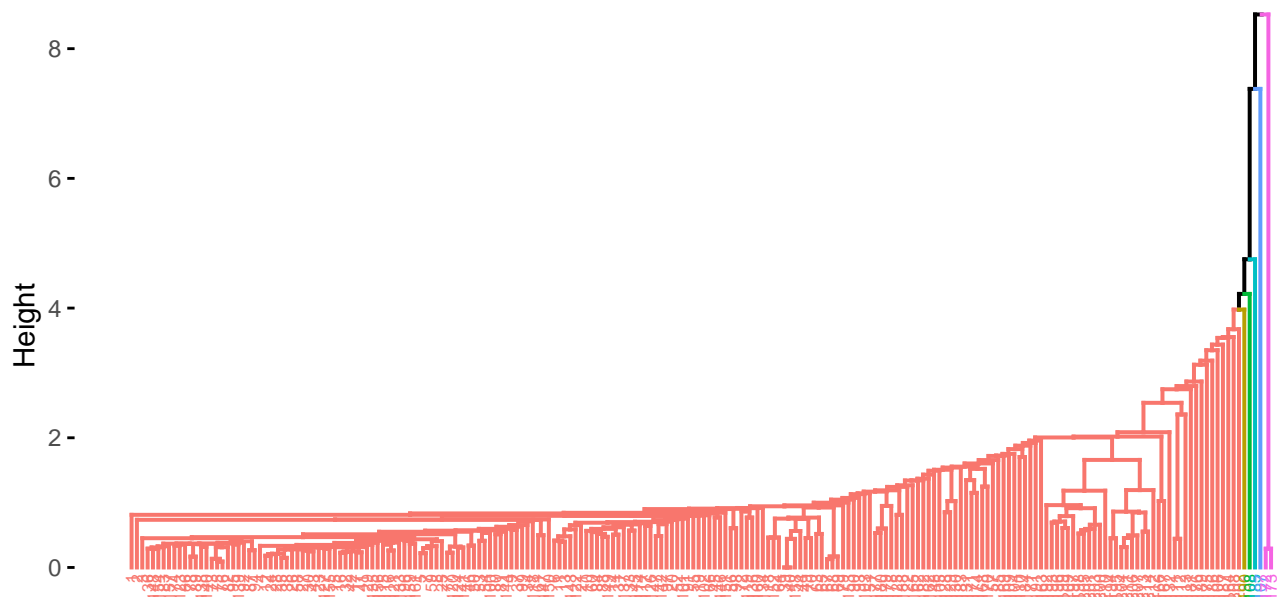
	1	2	3	5	6	7
1	23	7	4	0	5	2
2	47	61	13	2	0	1
3	0	0	0	2	0	0
5	0	7	0	5	0	0
6	0	1	0	3	4	13
7	0	0	0	1	0	13

Dokładność: 49.53%.

Gorzej niż k-średnie *shocked emoji*.

2.2.3 Agglomerative Nesting (AGNES)

Cluster Dendrogram



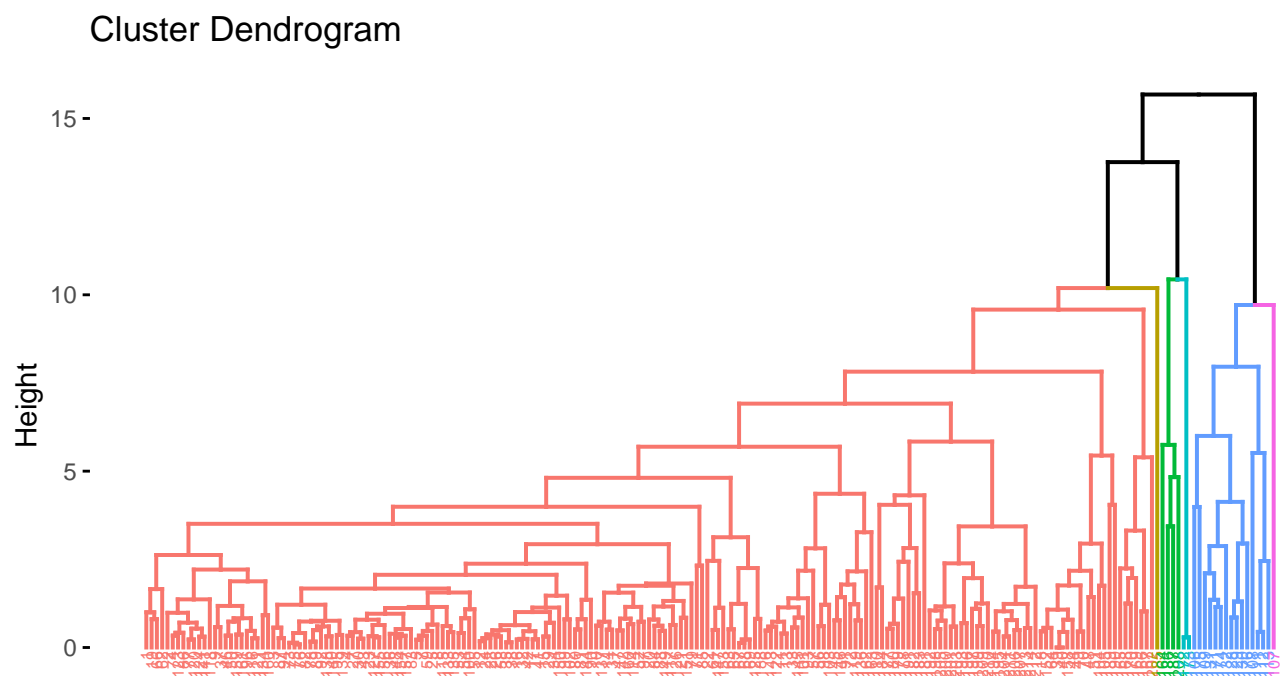
Wykres 8: AGNES: single linkage

2.2.3.1 Najbliższy sąsiad

Tabela 25: Macierz błędów; agnes, najbliższy sąsiad

	1	2	3	5	6	7
1	70	74	17	11	8	28
2	0	1	0	0	0	0
3	0	1	0	0	0	0
5	0	0	0	2	0	0
6	0	0	0	0	1	0
7	0	0	0	0	0	1

Dokładność: 35.05%.



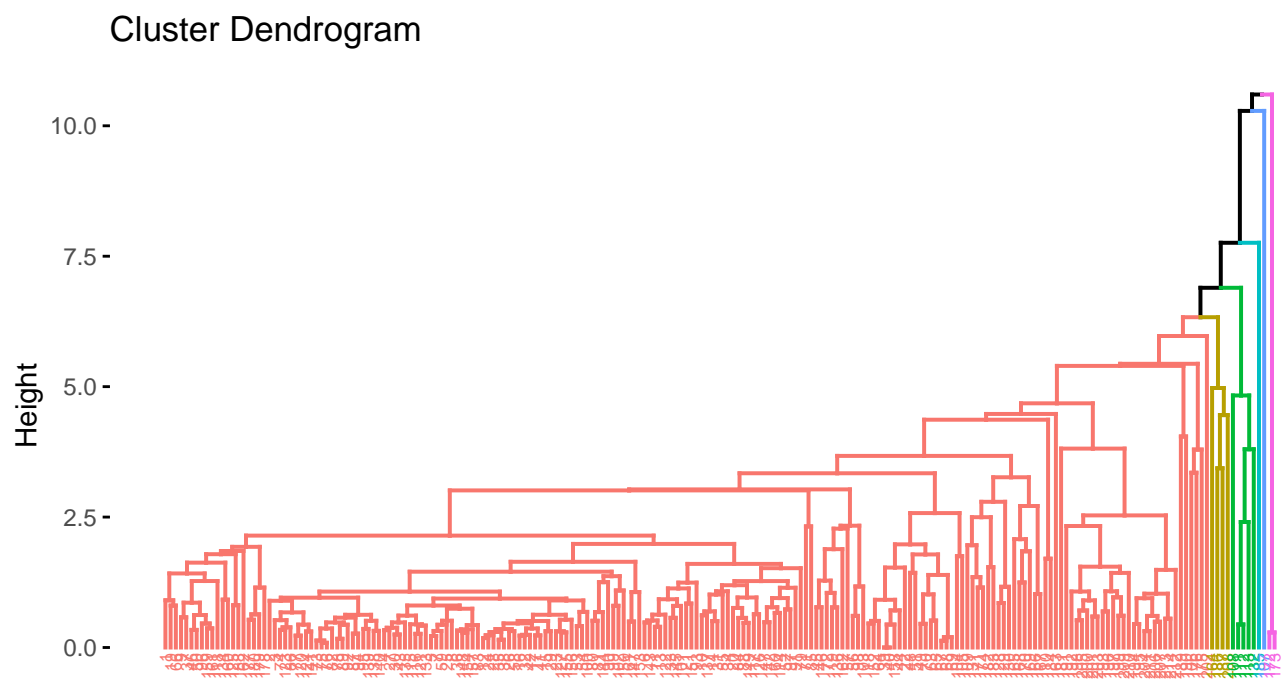
Wykres 9: AGNES: complete linkage

2.2.3.2 Najdalszy sąsiad

Tabela 26: Macierz błędów; agnes, najdalszy sąsiad

	1	2	3	5	6	7
1	70	64	17	6	8	26
2	0	11	0	4	0	0
3	0	1	0	0	0	0
5	0	0	0	1	0	3
6	0	0	0	2	0	0
7	0	0	0	0	1	0

Dokładność: 38.32%.



Wykres 10: AGNES: average linkage

2.2.3.3 Średnia odległość

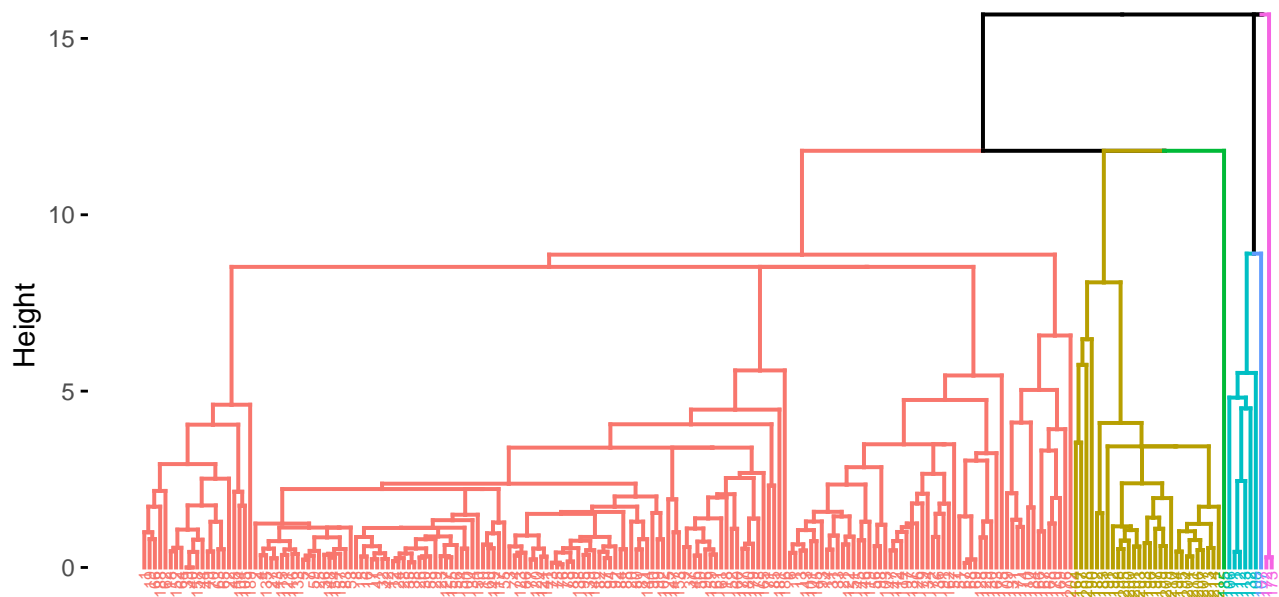
Tabela 27: Macierz błędów; agnes, średnia odległość

	1	2	3	5	6	7
1	70	70	17	10	8	26
2	0	1	0	0	0	0
3	0	5	0	0	0	0
5	0	0	0	1	0	3
6	0	0	0	2	0	0
7	0	0	0	0	1	0

Dokładność: 33.64%.

2.2.4 Divisive clustering (DIANA)

Cluster Dendrogram



Wykres 11: AGNES: average linkage

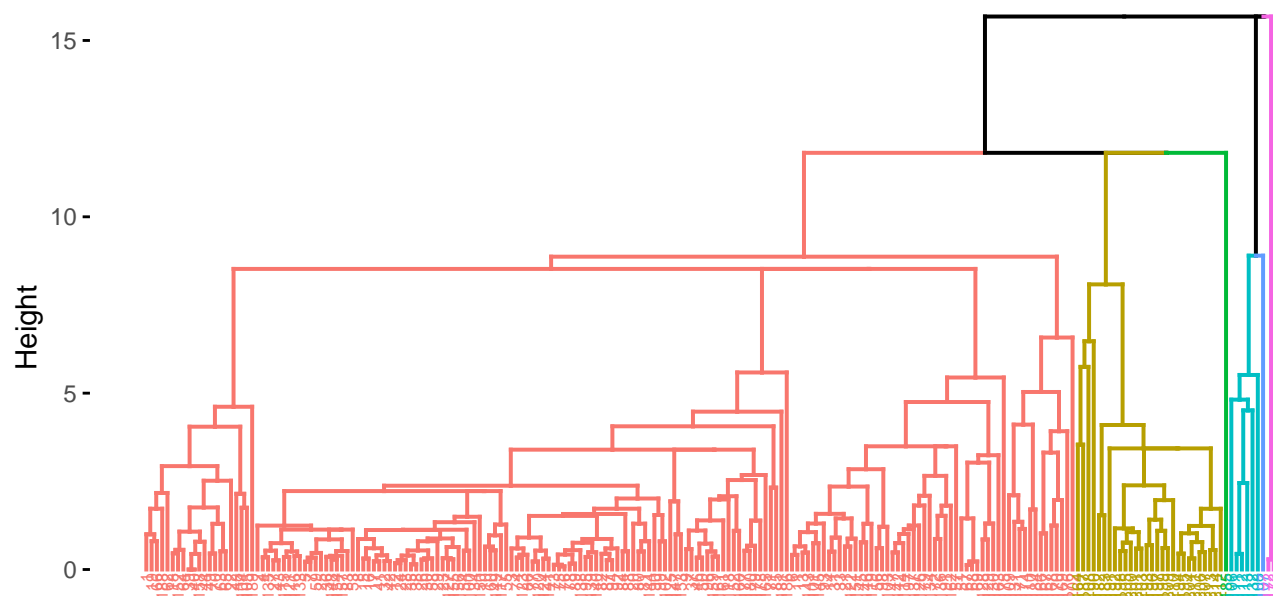
2.2.4.1 Odległość euklidesowa

Tabela 28: Macierz błędów; agnes, średnia odległość

	1	2	3	5	6	7
1	70	69	17	10	6	4
2	0	6	0	0	0	0
3	0	1	0	0	0	0
5	0	0	0	1	2	25
6	0	0	0	2	0	0
7	0	0	0	0	1	0

Dokładność: 35.98%.

Cluster Dendrogram



Wykres 12: AGNES: average linkage

2.2.4.2 Odległość Manhattan (taksówkowa)

Tabela 29: Macierz błędów; agnes, średnia odległość

	1	2	3	5	6	7
1	70	69	17	10	6	4
2	0	6	0	0	0	0
3	0	1	0	0	0	0
5	0	0	0	1	2	25
6	0	0	0	2	0	0
7	0	0	0	0	1	0

Dokładność: 35.98%.