

Sprawozdanie z listy 2

Eksploracja danych

Marta Stankiewicz (282244)

Paweł Nowak (282223)

2025-04-04

Spis treści

1	Zadanie 1.	1
1.1	Ocena zdolności dyskryminacyjnych zmiennych ciągłych.	1
1.2	Porównanie różnych metod dyskretyzacji nienadzorowanej.	2

Spis rysunków

1	Wykresy skrzypcowo-pudełkowe dla zmiennych ciągłych	2
---	---	---

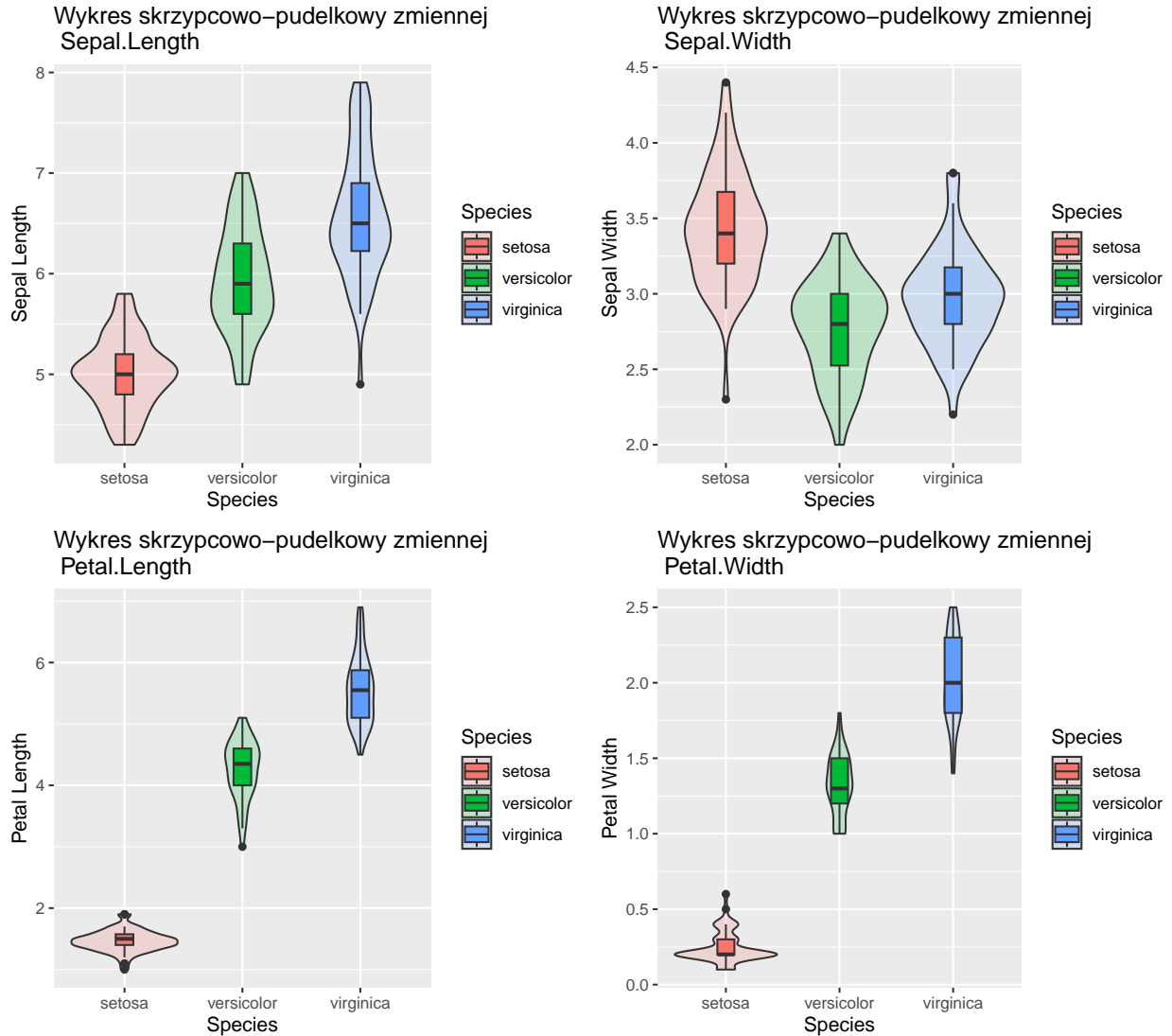
Spis tabel

1	Skuteczność wybranych metod dyskretyzacji dla zmiennej Sepal.Width . . .	2
2	Skuteczność wybranych metod dyskretyzacji dla zmiennej Petal.Width . . .	3

1 Zadanie 1.

1.1 Ocena zdolności dyskryminacyjnych zmiennych ciągłych.

W celu zbadania zdolności dyskryminacyjnej cech, posłużymy się wykresem skrzypcowo-pudełkowym (tj. wykresem skrzypcowym wraz z wykresem pudełkowym).



Rysunek 1: Wykresy skrzypcowo-pudełkowe dla zmiennych ciągłych

Z wykresów 1 wnioskujemy, że największe zdolności dyskryminacyjne wykazuje zmienna *Petal.Width*. Z kolei najmniejsze zdolności do separacji gatunków obserwujemy u zmiennej *Sepal.Width*.

1.2 Porównanie różnych metod dyskretyzacji nienadzorowanej.

Tabela 1: Skuteczność wybranych metod dyskretyzacji dla zmiennej *Sepal.Width*

equal_frequency	equal_width	cluster_based	fixed-bounds
0.91	0.81	0.91	0.73

Tabela 2: Skuteczność wybranych metod dyskretyzacji
dla zmiennej Petal.Width

equal_frequency	equal_width	cluster_based	fixed-bounds
0.97	1	0.99	0.86