

Komputerowa Analiza Danych

Zadanie 3

Cel:

1. Zaprezentowanie wyników grupowania danych przy pomocy algorytmu kMeans dla klasycznego zbioru irysów.
2. Zaprezentowanie wyników klasyfikacji danych przy pomocy algorytmu k-NN (najbliższych sąsiadów) dla klasycznego zbioru irysów.

Wynik:

Do wykonania tego zadania został użyty język programowania python. W celu wykonania zadania skorzystano z 4 bibliotek. Jedna służy do wczytania danych z pliku csv, druga do wykonywania obliczeń matematycznych, trzecia numpy do stworzenia wektora cech oraz matplotlib gdzie użyto subplots by móc umieścić wiele histogramów w jednym oknie aplikacji.

Wszystkie obliczenia zostały wykonane osobiście i znajdują się w 2 osobnych plikach kSrednich oraz KNN.

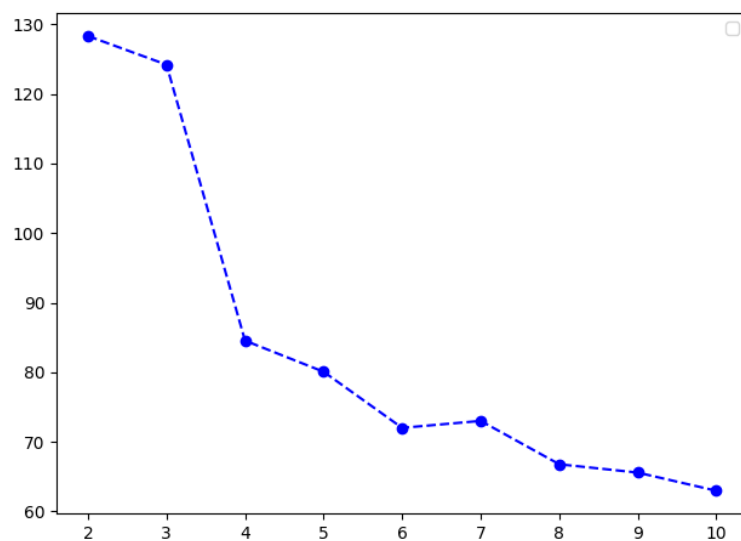
kMeans

Tabela 1 z liczbą iteracji w zależności od parametru k:

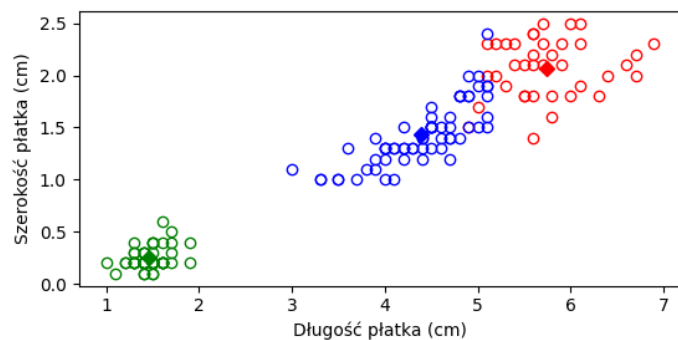
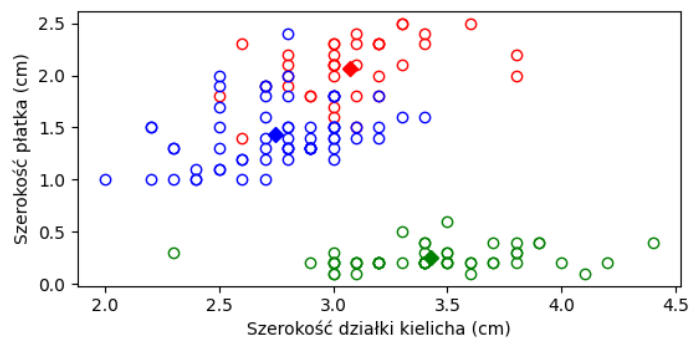
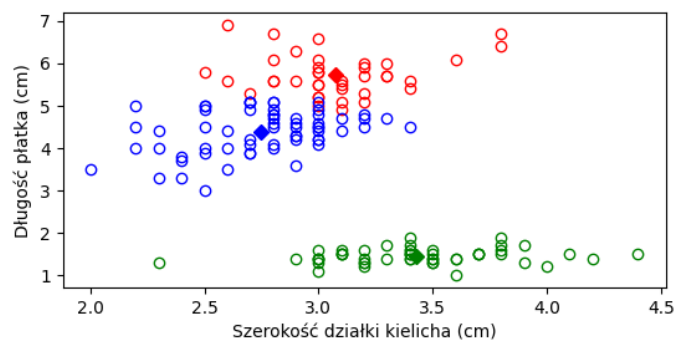
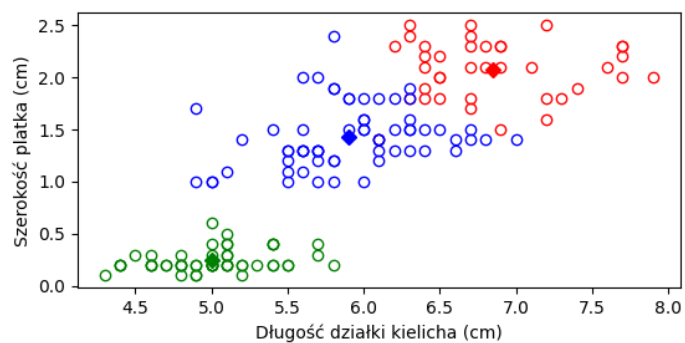
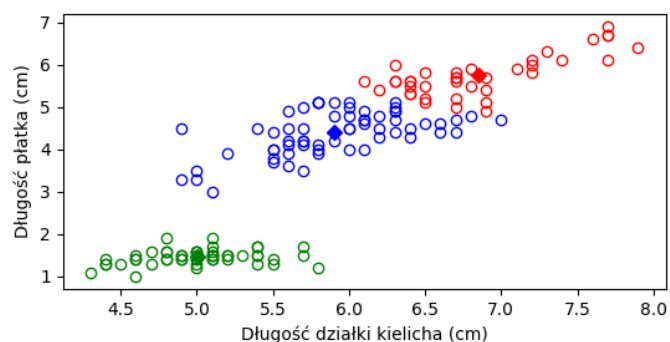
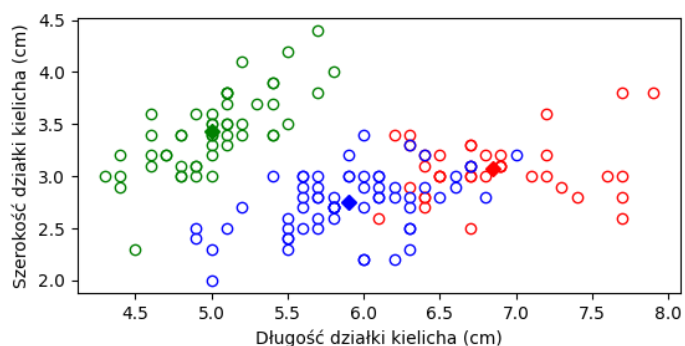
k	Ilość iteracji
2	15
3	14
4	15
5	20
6	21
7	22
8	23
9	33
10	37

Wykres 1 przedstawiający sumę kwadratów odległości WCSS dla grupowań z zakresu

$k = \langle 2, 10 \rangle$:



Wykres 2 przedstawiający grupowanie par cech, które zostały pogrupowane dla liczby klastrow równej 3:



Na tych wykresach jesteśmy w stanie zauważyć, że grupowanie to daje bardzo dobre rezultaty lecz niestety w niektórych parach np. szerokość działki kielicha oraz długość działki kielicha widzimy że dane mieszają się. Jest to spowodowane tym, że próbujemy przedstawić 4 wymiarowy wektor na dwuwymiarowych wykresach par.

k-NN

W tym wypadku wykorzystałem podobnie jak w Zadaniu 1 biblioteki tabulate do schludnego przedstawienia wyników macierzy.

Wykres 3 przedstawiający procentowy wynik klasyfikacji zbioru testowego dla 4 parametrów,

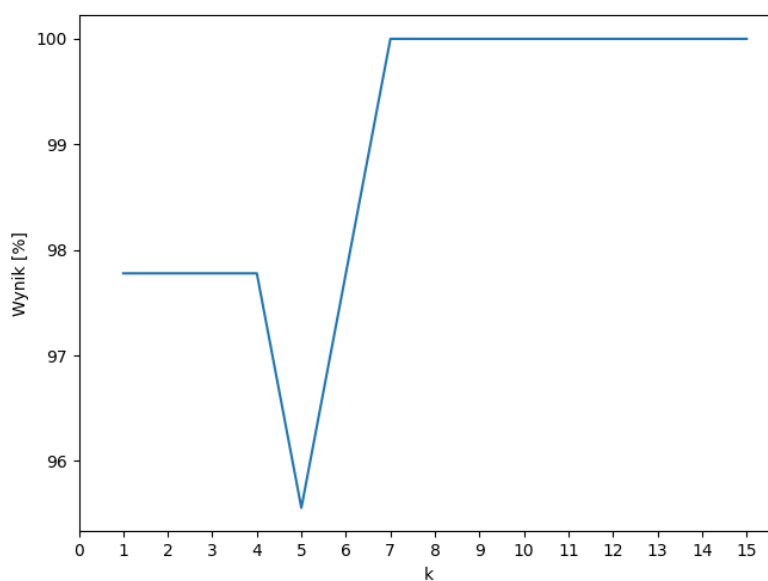


Tabela 2 przedstawia macierz pomyłek dla 4 parametrów.

k = 7 Rozpoznanie Faktyczna klasa v	Setosa	Versicolor	Virginica
Setosa	15	0	0
Versicolor	0	15	0
Virginica	0	0	15

Wykres 4 przedstawia wyniki klasyfikacji Długość Działki Kielicha(cm) i Szerokość Działki Kielicha(cm)

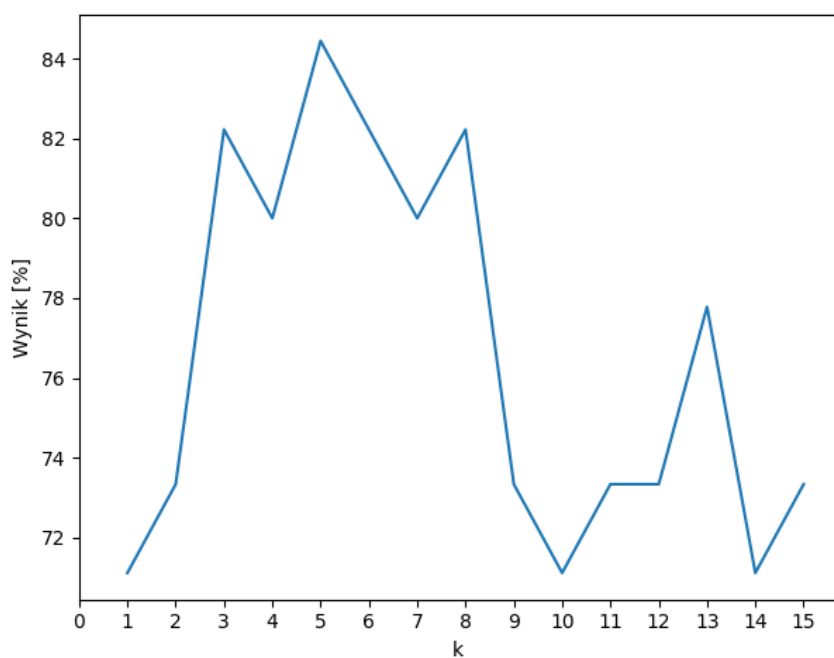


Tabela 3 przedstawia macierz pomyłek dla Długość Działki Kielicha(cm) i Szerokość Działki Kielicha(cm)

k = 5 Rozpoznanie Faktyczna klasa v	Setosa	Versicolor	Virginica
Setosa	15	0	0
Versicolor	0	8	7
Virginica	0	5	10

Wykres 5 przedstawia wyniki klasyfikacji Długość Działki Kielicha(cm) i Długość Płatka(cm)

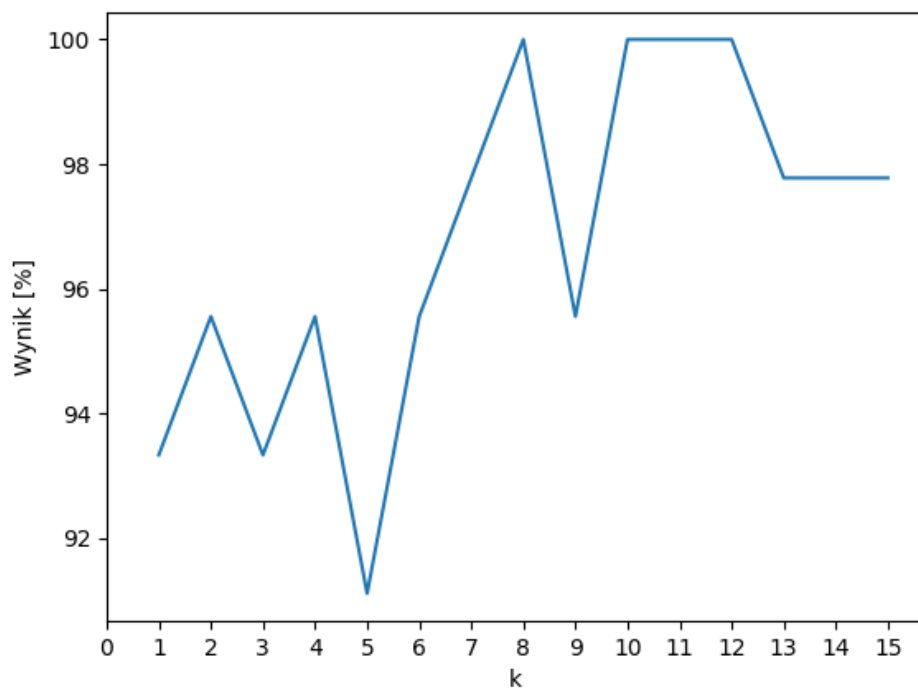


Tabela 4 przedstawia macierz pomyłek dla Długość Działki Kielicha(cm) i Długość Płatka(cm)

k = 8 Rozpoznanie Faktyczna klasa v	Setosa	Versicolor	Virginica
Setosa	15	0	0
Versicolor	0	14	1
Virginica	0	0	15

Wykres 6 przedstawia wyniki klasyfikacji dla Długość Działki Kielicha(cm) i Szerokość Płatka(cm)

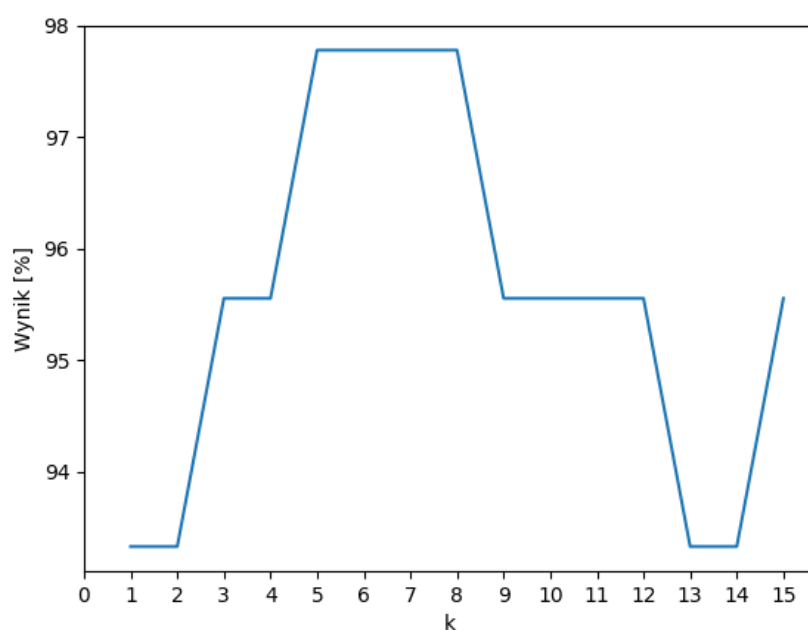


Tabela 5 przedstawia macierz pomyłek dla Długość Działki Kielicha(cm) i Szerokość Płatka(cm)

k = 5 Rozpoznanie Faktyczna klasa v	Setosa	Versicolor	Virginica
Setosa	15	0	0
Versicolor	0	13	2
Virginica	0	0	15

Wykres 7 przedstawia wyniki klasyfikacji dla Szerokość Działki Kielicha(cm) i Długość Płatka(cm)

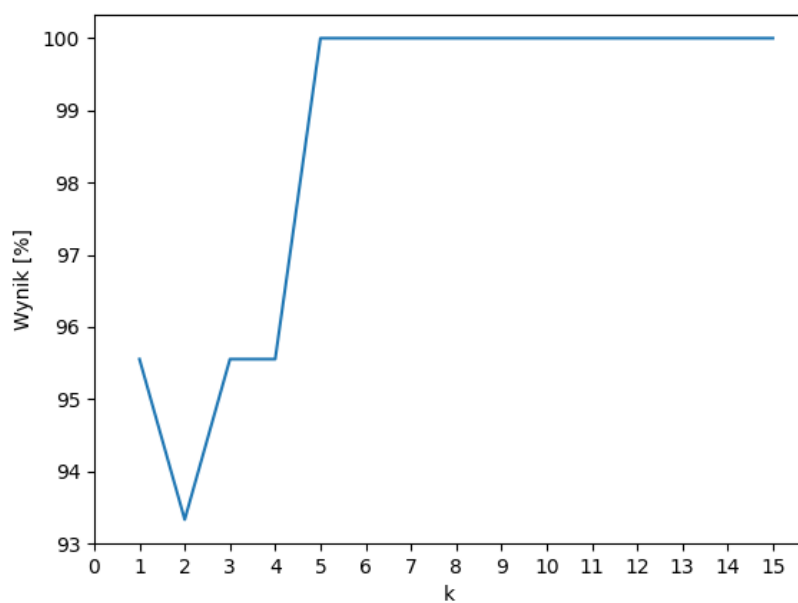


Tabela 6 przedstawia macierz pomyłek dla Szerokość Działki Kielicha(cm) i Długość Płatka(cm)

k = 5 Rozpoznanie Faktyczna klasa v	Setosa	Versicolor	Virginica
Setosa	15	0	0
Versicolor	0	15	0
Virginica	0	0	15

Wykres 8 przedstawia wyniki klasyfikacji dla Szerokość Działki Kielicha(cm) i Szerokość Płatka(cm)

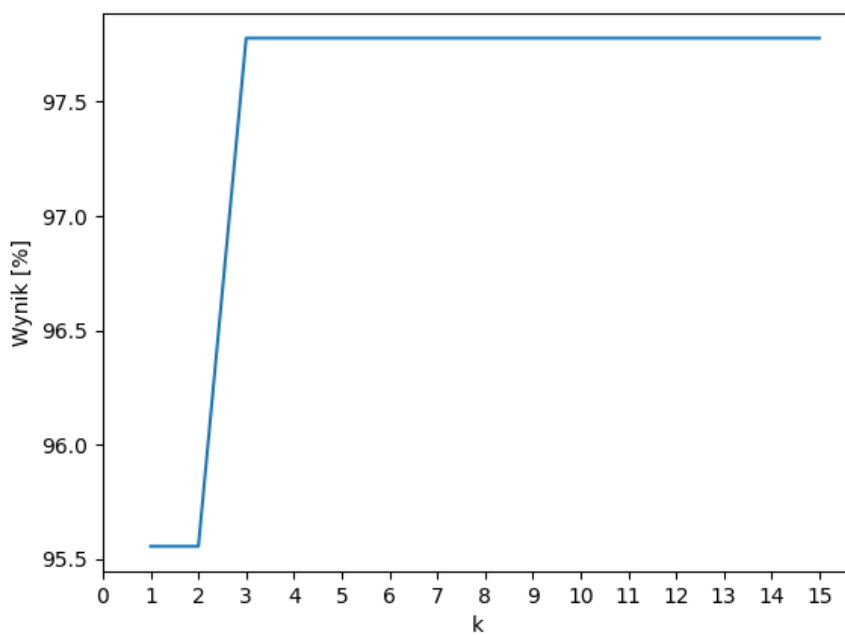


Tabela 7 przedstawia macierz pomyłek dla Szerokość Działki Kielicha(cm) i Szerokość Płatka(cm)

k = 3 Rozpoznanie Faktyczna klasa v	Setosa	Versicolor	Virginica
Setosa	15	0	0
Versicolor	0	15	0
Virginica	0	1	14

Wykres 9 przedstawia wyniki dla Długość Płatka(cm) i Szerokość Płatka(cm)

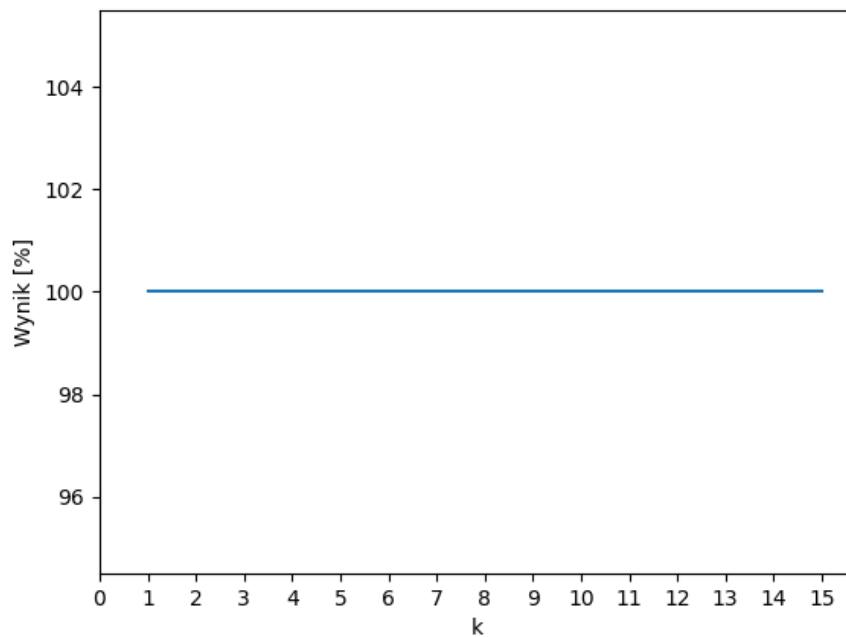


Tabela 8 przedstawia macierz pomyłek dla Długość Płatka(cm) i Szerokość Płatka(cm)

k = 1 Rozpoznanie Faktyczna klasa v	Setosa	Versicolor	Virginica
Setosa	15	0	0
Versicolor	0	15	0
Virginica	0	0	15