SPRAWOZDANIE

Zajęcia: Uczenie Maszynowe

Prowadzący: prof. dr hab. Vasyl Martsenyuk

|  |  |
| --- | --- |
| Laboratorium Nr 3  Data 07.12.2024  Temat: "Uczenie maszynowe w praktyce: analiza skupień"  Wariant 10 | Anna Więzik  Informatyka  II stopień, niestacjonarne,  1 semestr, gr.1b |

1. Polecenie:

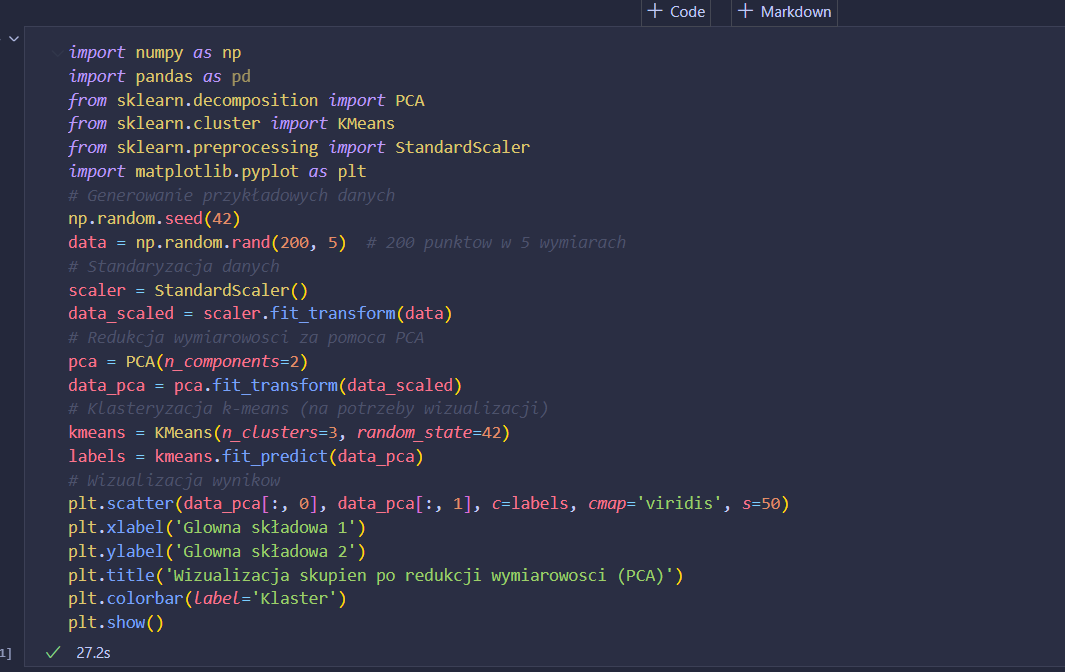
Smoking patient: <https://www.kaggle.com/datasets/thomaskonstantin/cpg-values-of-smoking-and-non-smoking-patients>

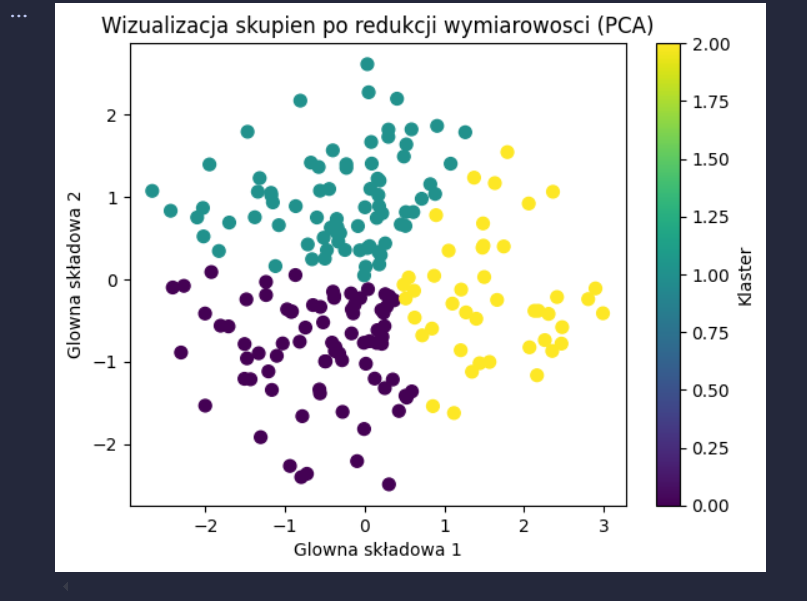
1. Link do repozytorium:

Link: <https://github.com/AnaShiro/UM_2024>

1. Opis programu opracowanego
2. Wizualizacja skupień

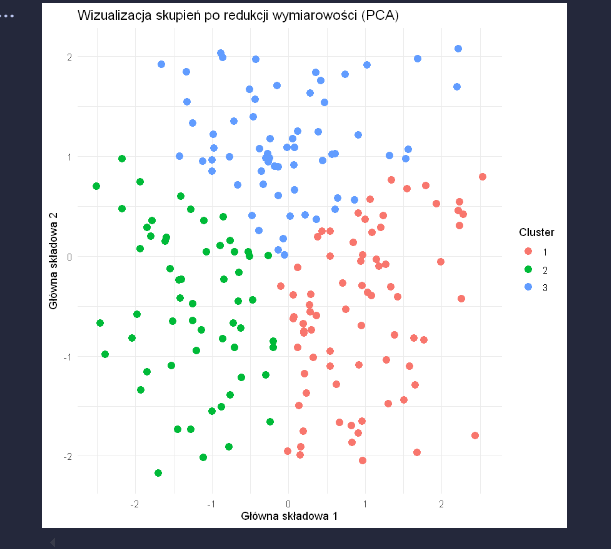
* Python





* R

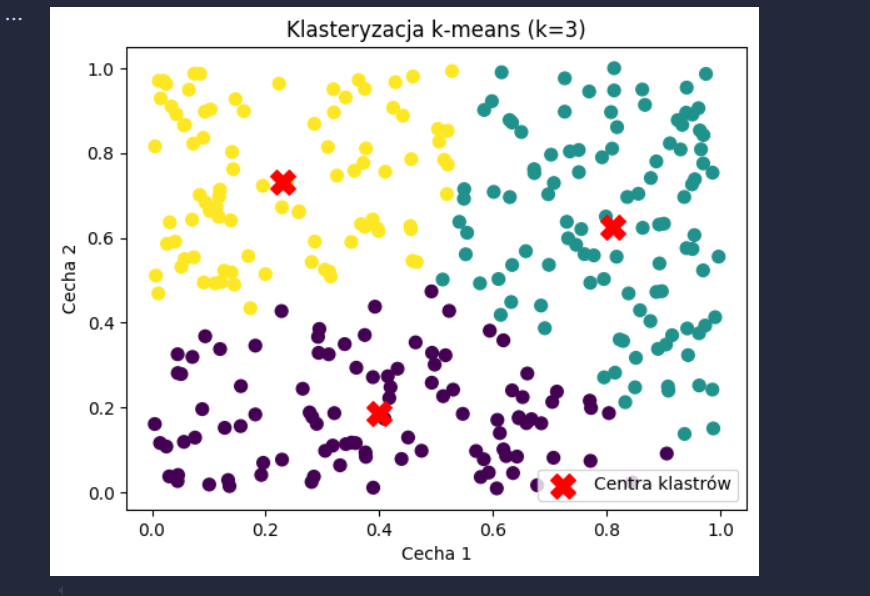




1. Klasteryzacja k-means

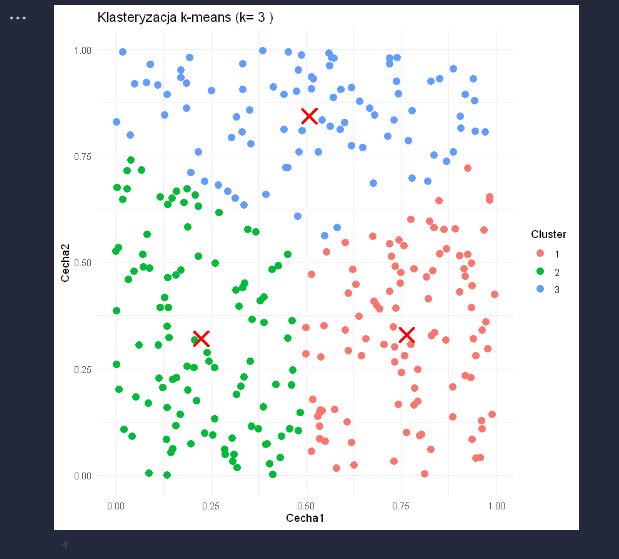
* Python





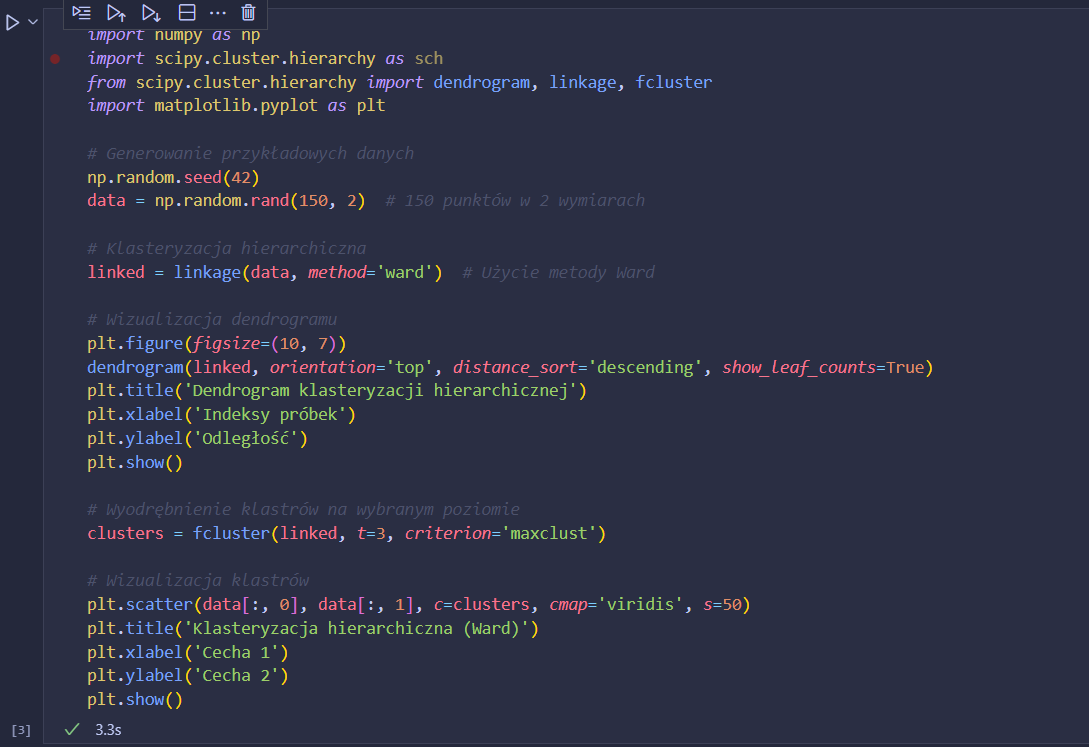
* R

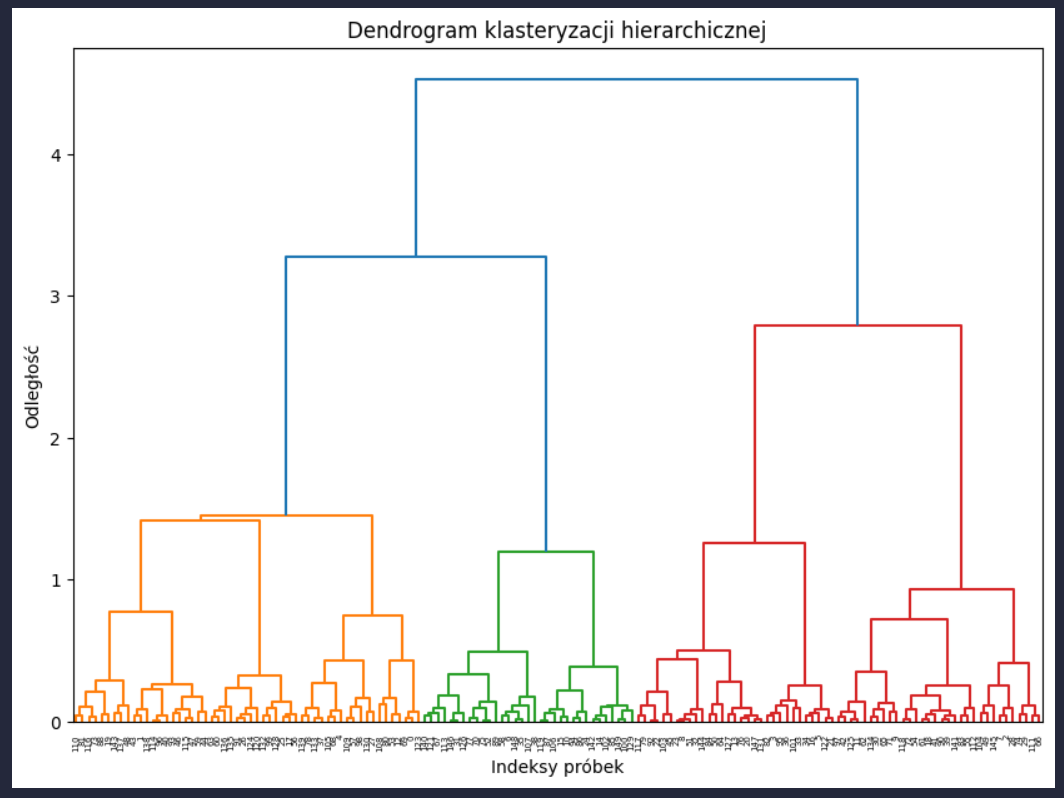


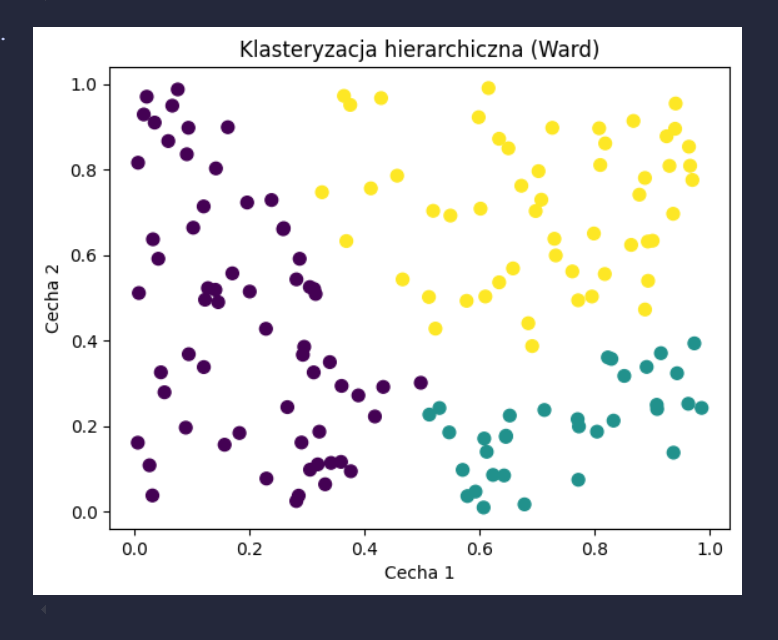


1. Klasteryzacja hierarchiczna

* Python

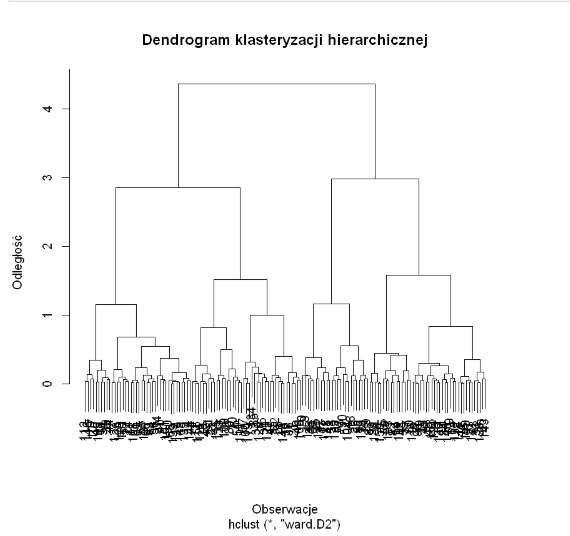


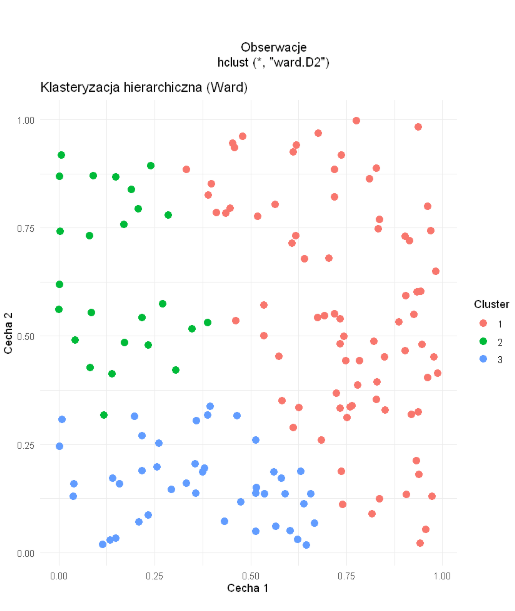




* R







1. Wnioski

Uczenie nienadzorowane jest jednym z kluczowych typów uczenia maszynowego, w którym odkrywa się ukryte wzorce w danych bez potrzeby posiadania etykiet (czyli zmiennych wynikowych). Analiza skupień (ang. clustering) to proces grupowania danych w k klastrów, gdzie punkty w ramach jednego klastra są bardziej podobne do siebie niż do punktów z innych klastrów. Celem klasteryzacji jest minimalizacja funkcji kosztu J(C), która mierzy wewnętrzną różnorodność w klastrach. Ocena wyników klasteryzacji w uczeniu nienadzorowanym jest bardziej skomplikowana niż w przypadku uczenia nadzorowanego, ponieważ brak jest etykiet referencyjnych. Analiza głównych składowych (ang. Principal Component Analysis, PCA) to metoda redukcji wymiarowości danych, mająca na celu przekształcenie danych wielowymiarowych w mniejszą liczbę wymiarów przy zachowaniu maksymalnej ilości informacji. Metody niehierarchiczne do analizy skupień, takie jak k-means, polegają na podziale danych na k skupień, gdzie k jest parametrem wybranym przez użytkownika. Metody hierarchiczne budują hierarchię klastrów, umożliwiając zrozumienie struktury danych na różnych poziomach szczegółowości. Metryka odległości (np. odległość euklidesowa, Manhattan, kosinusowa) określa sposób mierzenia odległości między punktami, natomiast metoda łączenia (np. single linkage, complete linkage, average linkage) określa, jak mierzyć odległość między klastrami.