

ED 6

Zadanie 1

- A. Wczytaj dane z pliku [k_means_data.csv](#).
- B. Zastosuj **algorytm hierarchiczny aglomeracyjny** i wyszukaj w zbiorze danych klastry.
- C. Zastosuj **algorytm k-średnich** i wyszukaj w zbiorze danych klastry.

Zadanie 2

Zaimplementuj samodzielnie algorytm k-średnich dla danych z pliku [k_means_data.csv](#).

Zadanie 3

Napisz program pozwalający obliczać podobieństwo między sekwencjami DNA traktowanymi jako ciągi znaków.

Zadanie 4

1. Załaduj do pliku zbiór danych [iris.csv](#).
Opis zbioru: https://scikit-learn.org/stable/auto_examples/datasets/plot_iris_dataset.html



2. Narysuj **wykres rozrzutu** dla współrzędnych **sepal_length** i **sepal_width**. Spróbuj oznaczyć kolorami różne klasy kwiatów.
3. Znajdź wartości **minimalne**, **maksymalne** oraz **średnie odchylenie kwadratowe** dla atrybutów zbioru.
4. Policz **współczynniki korelacji** między atrybutami.
5. Czy prawdziwe jest stwierdzenie: **petal_length** klasy **Setosa** jest krótsze niż **petal_length** innych klas?
6. Zastosuj **algorytm k-średnich** i znajdź **podział na klastry** dla współrzędnych **sepal_length** i **sepal_width**.
7. Zastosuj **algorytm hierarchiczny aglomeracyjny** i znajdź **podział na klastry** dla współrzędnych **sepal_length** i **sepal_width**.