### 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

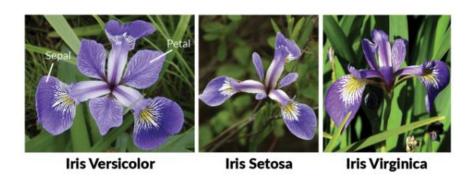
Zaimplementuj samodzielnie algorytm k-średnich dla danych z pliku k\_means\_data.csv.

### Zadanie 3

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

## Zadanie 4

Załaduj do pliku zbiór danych iris.csv.
Opis zbioru: <a href="https://scikit-learn.org/stable/auto">https://scikit-learn.org/stable/auto</a> 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** <u>jest krótsze niż</u> **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**.