

Metody i narzędzia BIG DATA

Lista 6

Zad. 1 Korzystając z bibliotek dostępnych w języku Python rozwiąż zadanie klasyfikacji wykorzystując algorytmy:

1. k-Najbliższych Sąsiadów (ang. *k-Nearest Neighbors*)
2. Maszyna wektorów nośnych (ang. *Support Vector Machines*)
3. Drzewa decyzyjne (ang. *Decision Tree*)
4. Inny, wybrany samodzielnie

oraz następujące zbiory dostępne pod adresem <https://scikit-learn.org/stable/datasets.html>

1. Iris plants dataset
2. Wine recognition dataset
3. Inny, wybrany samodzielnie

Przeprowadź analizę jakości otrzymanych wyników przez każdy z klasyfikatorów oraz porównaj wyniki otrzymane przez te klasyfikatory.

Do oceny jakości klasyfikacji należy wykorzystać wybrane metody

1. Tablicę pomyłek
2. Miary jakości takie jak dokładność, czułość, precyzję, f-1
3. Krzywa ROC

Zad. 2 Korzystając z bibliotek dostępnych w języku Python rozwiąż zadanie grupowania wykorzystując algorytmy:

1. k-średnich (ang. *k-Means*)
2. Przesunięcie średniej (ang. *Mean Shift*)
3. Inny, wybrany samodzielnie

oraz dwa wybrane zbiory danych dostępne pod adresami https://scikit-learn.org/stable/datasets/sample_generators.html

Do oceny jakości grupowania należy wykorzystać wybrane metody

1. Współczynnik kształtu (ang. *Silhouette Coefficient*)
2. Indeks Calińskiego-Harabasa (ang. *Calinski-Harabasz Index*)
3. Indeks Randa (ang. *Rand index*)
4. Jednorodność (ang. *Homogeneity*)
5. Informacja wzajemna (ang. *Mutual Information*)

Dokonaj analizy porównawczej algorytmów. Wykonaj odpowiednie przedstawienia graficzne uzyskanych wyników.

Zad. 3 Sformułuj oraz rozwiąż zadanie klasyfikacji dla samodzielnie wybranego zbioru danych.