

WSI - ćwiczenie 7.

Modele bayesowskie

10 stycznia 2023

1 Sprawy organizacyjne

1. Ćwiczenie realizowane jest samodzielnie.
2. Ćwiczenie wykonywane jest w języku Python.
3. Ćwiczenie powinno zostać oddane najpóźniej na ostatnich zajęciach. W ramach oddawania ćwiczenia należy zademonstrować prowadzącemu działanie kodu oraz utworzyć pull request (z kodem oraz raportem) który prowadzący będzie mógł komentować.
4. Rozwiązanie ćwiczenia powinno być zgodne z szablonem z repozytorium <https://gitlab-stud.elka.pw.edu.pl/jlyskawa/wsi-template>.
5. Dokumentacja powinna być w postaci pliku .pdf, .html albo być częścią notebooka jupyterowego. Powinna zawierać opis eksperymentów, uzyskane wyniki wraz z komentarzem oraz wnioski.
6. Na ocenę wpływa poprawność oraz jakość kodu i dokumentacja.
7. Można korzystać z pakietów do obliczeń numerycznych, takich jak *numpy*
8. Można skorzystać z pakietu *scikit-learn* w celu załadowania zbioru danych oraz wykonania walidacji krzyżowej
9. Implementacja powinna być ogólna.

2 Ćwiczenie

Celem ćwiczenia jest implementacja naiwnego klasyfikatora Bayesa.

Następnie należy wykorzystać stworzony algorytm do stworzenia i zbadania jakości klasyfikatorów dla zbioru danych Cardio Vascular Disease Detection (<https://www.kaggle.com/datasets/bhadaneeraj/cardio-vascular-disease-detection>). Klasą jest pole *cardio*.

Należy sprawdzić jakość zarówno różnymi podziałami na zbiór trenujący, walidacyjny i testowy jak i różnymi uruchomieniami k-krotnej walidacji krzyżowej (z różnymi podziałami na zbiór treningowy+walidacyjny i testowy). Należy zbadać, która metoda pomiaru (zbiór walidacyjny czy walidacja krzyżowa) lepiej odzwierciedla wyniki na zbiorach testowych.