

WSI - ćwiczenie 4.

Klasyfikacja

14 kwietnia 2025

1 Sprawy organizacyjne

1. Ćwiczenie realizowane jest samodzielnie.
2. Ćwiczenie wykonywane jest w języku Python.
3. Ćwiczenie powinno zostać oddane najpóźniej na 9. zajęciach. W ramach oddawania ćwiczenia należy zademonstrować prowadzącemu działanie kodu oraz wysłać na maila kod oraz dokumentację.
4. Dokumentacja powinna być w postaci pliku .pdf, .html albo być częścią notebooka jupyterowego. Powinna zawierać opis eksperymentów, uzyskane wyniki wraz z komentarzem oraz wnioski.
5. Na ocenę wpływa poprawność oraz jakość kodu i dokumentacja.
6. Można korzystać z pakietów do obliczeń numerycznych, takich jak *numpy*

2 Ćwiczenie

Zaimplementować algorytm lasu losowego oraz przetestować działanie dla zadania klasyfikacji na zbiorze danych zawierającym informacje o wystąpieniu chorób serca (poprawną predykcję wartości z kolumny HeartDisease). Rozwiązanie powinno zawierać elementy:

1. Opis działań związanych przygotowaniem zbioru danych pod zagadnienie klasyfikacji.
2. Obliczenie wartości metryk: dokładność (ang. *accuracy*), precyzja (ang. *precision*), czułość (ang. *recall*), krzywa ROC (ang. *ROC curve*) oraz pole pod krzywą (można korzystać z gotowych rozwiązań dla krzywej ROC, np. Scikit-learn).
3. Próba poprawy wartości metryk za pomocą optymalizacji hiperparametrów: liczby drzew.
4. Gdyby podany model był używany w rzeczywistości (np. w ochronie zdrowia) do przewidywania chorób serca, która z metryk powinna być maksymalizowana i dlaczego?

Wskazówka: W Internecie można znaleźć dość dużo frameworków zawierających implementacje klasyfikatorów i modeli regresyjnych w tym Lasu Losowego (np. Scikit-learn dla Python'a). Jest to dość dobra (a przede wszystkim szybka) metoda na walidację wyników uzyskanych dla własnej implementacji.