

WSI-24L-G104

Adam Kwaśnik 10.04.2024

Zadanie 4.

Cel zadania

Celem zadania jest wykorzystanie modelu SVM oraz Drzew decyzyjnych w klasyfikacji danych. Dataset, na którym będą Państwo pracować znajduje się [w tym linku](#). Zawiera on informacje na temat win i ich przynależności do konkretnej klasy.

W trakcie implementacji algorytmów polecam wykorzystanie biblioteki scikit-learn.

W celu wykonania zadania, muszą Państwo:

- pobrać, odpowiednio wczytać dane.
- wytrenować modele korzystając z walidacji krzyżowej (należy podzielić zestaw danych na 4 części. Każda część w kolejnej iteracji uczenia będzie pełnić rolę zbioru testowego)
- Dokonać oceny skuteczności wytrenowanego modelu korzystając z następujących metryk: Accuracy, Precision, Recall, F1 Score¹

¹ [metryki info](#)

Implementując algorytm SVM oraz drzewa decyzyjne możemy zmieniać wiele parametrów. Dla każdego z nich należy przetestować ich wpływ na wyniki:

SVM:

- Funkcja jądra (kernel)
- Siła regularyzacji
- Liczba iteracji (przy analizie wpływu tego parametru trzeba znacząco zmniejszyć parametr tol, którego domyślna wartość to 1e-3. Parametr tol służy jako kryterium zatrzymania procesu uczenia algorytmu)

Drzewa decyzyjne:

- Maksymalna głębokość drzewa
- Kryterium używane do oceny jakości
- Strategia podziału węzła

Implementacja tych algorytmów w bibliotece scikit-learn zawiera również parametr `random_state` odpowiadający za kontrolowanie losowości. Wszystkie pomiary należy powtórzyć 3 krotnie z różnym parametrem `random_state` aby mieć 3 oszacowania skuteczności każdego modelu.

Sprawozdanie

W sprawozdaniu należy umieścić opis działania algorytmów i zestawienie wyników skuteczności algorytmów. Proszę również o dołączenie wykresów przedstawiających jak zmienia się skuteczność modelu w zależności od zmiany wybranego parametru liczbowego takiego jak np. liczba iteracji w przypadku SVM.