

Wydział Elektroniki i Technik Informacyjnych
Politechnika Warszawska

Wprowadzenie do sztucznej inteligencji

Sprawozdanie z ćwiczenia nr 4

Tymon Kobylecki

Warszawa, 2022

Spis treści

1. Wstęp	2
2. Ćwiczenie	3
2.1. Eksperymenty	3
2.2. Środowisko - dane o raku piersi	3
2.3. Wyniki	3
2.4. Analiza wyników	3
2.5. Wnioski	4

1. Wstęp

W niniejszym sprawozdaniu opisane zostało rozwiązanie zadania oraz eksperymenty dotyczące zadania nr 3 polegającego na implementacji drzew decyzyjnych tworzonych algorytmem ID3 z ograniczeniem maksymalnej głębokości drzewa. Całość ćwiczenia została wykonana w języku Python. Wykorzystana została implementacja drzewa z biblioteki `anytree`.

2. Ćwiczenie

2.1. Eksperymenty

Eksperymenty w niniejszym ćwiczeniu laboratoryjnym polegały na 25-krotnym uruchamianiu programu, który uruchamiał algorytm ID3 dziesięciokrotnie i na tej podstawie znajdował optymalną głębokość dla wylosowanego zestawu danych. Dane były dzielone w proporcji 95:95:96.

2.2. Środowisko - dane o raku piersi

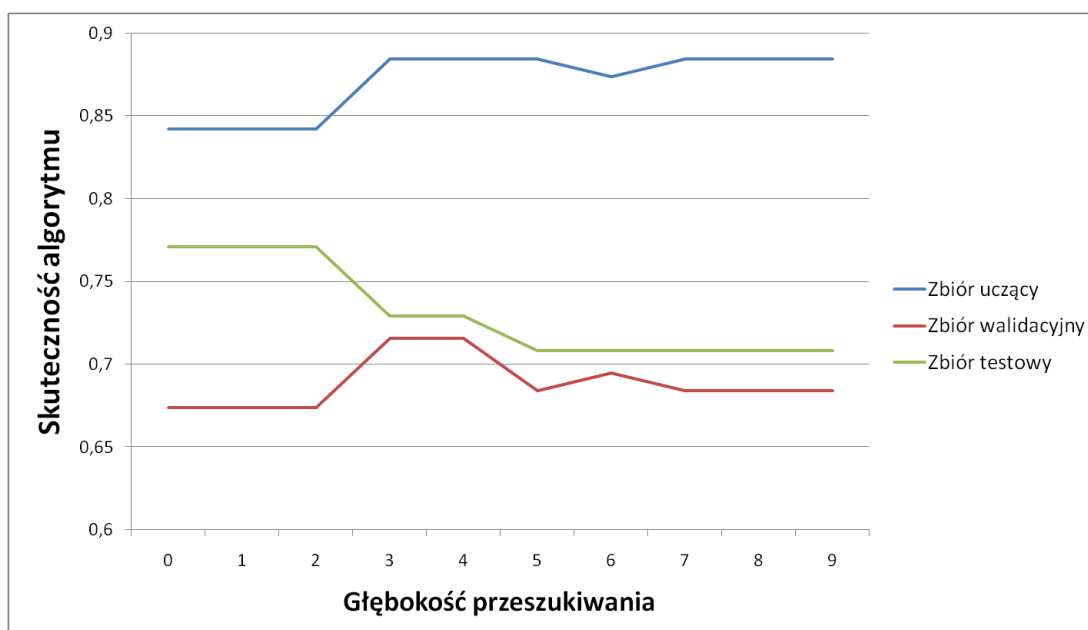
Danymi dostarczonymi algorytmowi były dane o raku piersi i cechach pacjentek. Dane zostały zaczerpnięte z <https://archive.ics.uci.edu/ml/datasets/breast+cancer>.

2.3. Wyniki

Na wykresie 2.1 zostały przedstawione uśrednione wyniki z 25 uruchomień programu, z których każde dostarczyło po jednej wartości dla każdej głębokości dla każdego zbioru.

2.4. Analiza wyników

Łatwo zauważyć, że między głębokością 3 a 4 na zbiorze walidacyjnym obserwujemy maksimum skuteczności algorytmu. Na dalszych odcinkach mamy bowiem do czynienia z przeuczeniem, co skutkuje obniżeniem skuteczności na wszystkich zbiorach oprócz uczącego.



Rys. 2.1. Wykres przedstawiający uśrednione wyniki dla wszystkich głębokości

2.5. Wnioski

Algorytm dość dobrze radzi sobie z postawionym zadaniem, jednak przy takim zestawie danych, gdzie zachodzi wyraźna dominacja jednego wyniku, może zachodzić tendencja do „upraszczania” zadania przez algorytm, czyli dobierania głębokości równej 0. Wówczas algorytm stwierdza, że opłaca mu się zawsze mówić, że pacjentka nie ma raka, gdyż zdecydowana większość pacjentek go nie ma.