

WSI 4

Adam Sokołowski 324892

May 2024

1 Cel ćwiczenia

Zaimplementować algorytm ID3 z ograniczeniami maksymalnej głębokości drzewa, oraz przetestować jego działanie przy użyciu zbioru danych Cardio Vascular Disease Detection. Należało również znaleźć najlepszą głębokość.

2 Realizacja, przeprowadzone eksperymenty oraz wnioski

Najpierw atrybuty które nie były dyskretne zdyskretyzowano. Dyskretne dane podzielono na części i zapisano w oddzielnych plikach. Skorzystano z gotowej metody train test split z biblioteki sklearn.

Aby przetestować działanie porównano działanie mojego algorytmu z gotowym od sklearn. Do porównania skorzystano z gotowej metody classification report z sklearn.metrics. Wykorzystano te same dane.

ID3 Decision Tree (My Implementation) Report:				
	precision	recall	f1-score	support
0	0.62	0.80	0.70	6935
1	0.72	0.51	0.60	6926
accuracy			0.66	13861
macro avg	0.67	0.66	0.65	13861
weighted avg	0.67	0.66	0.65	13861

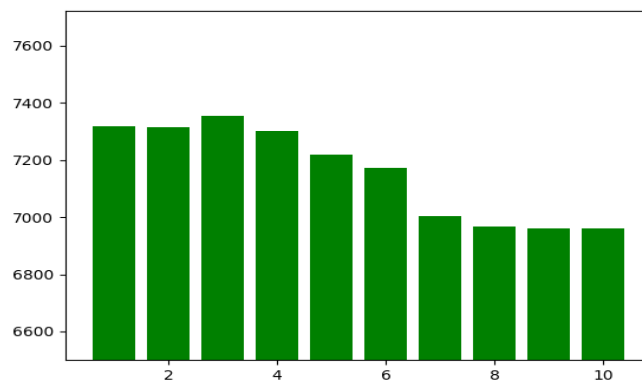
Rysunek 1: Raport z mojego algorytmu dla maksymalnej głębokości 5

Sklearn Decision Tree Report:				
	precision	recall	f1-score	support
0	0.69	0.81	0.75	6935
1	0.77	0.64	0.70	6926
accuracy			0.73	13861
macro avg	0.73	0.73	0.72	13861
weighted avg	0.73	0.73	0.72	13861

Rysunek 2: Raport z algorytmu od sklearn

Jak widać na powyższych rysunkach wyniki są bardzo podobne i na tej podstawie można stwierdzić, że mój algorytm działa poprawnie, jednak trochę gorzej od tego gotowego.

Następnie przeprowadzono badania najlepszej głębokości. Sprawdzono głębokości od 1 do 10 i zapisano liczbę zgadzających się przewidywań tego czy osoba jest chora czy nie w danej próbie (liczba rzędów danych w zbiorze testowym równa 11089). Z wykresu wynika, że najlepsza głębokość to 3.



Rysunek 3: Wykres zależności poprawności przewidywań od głębokości maksymalnej