

Jakub Kowalczyk Ćwiczenie 4. – Implementacja ID3

Celem eksperymentu było sprawdzenie wpływu hiperparametru głębokości drzewa na poprawność przewidzianych klas w drzewie decyzyjnym tworzonym algorytmem ID3. Eksperyment wymagał ponadto odpowiedniego zdyskretyzowania wartości ciągłych w zbiorze danych.

Decyzje projektowe

Dane z pliku csv wczytałem za pomocą biblioteki pandas, która zwraca je jako obiekt DataFrame i to na nim operowałem. Wyboru przedziałów zbiorów dla wartości ciągłych dokonałem, wykonując obliczenia w Excelu. Zdecydowałem się na podzielenie każdego niedyskretnego atrybutu na 3 przedziały, które miały zbliżoną między sobą liczebność. Następnie w Pythonie dokonałem mapowania wartości na liczby 1, 2 lub 3. Podział przedstawia się tak:

	Grupa 1	Grupa 2	Grupa 3
Wiek	<=50,0	50-57	57+
	21210	24610	23481
Wzrost	<=160	161-169	170+
	23283	28188	17830
Waga	<=67	68-79	80+
	23571	24277	21453
Ap_hi	<=119	120	121+
	12913	27431	28957
Ap_lo	<=79	80	81+
	13974	34513	20814

Tabela 1

Zgodnie z zaleceniami podzieliłem zbiór danych w następujących proporcjach:

Zbiór treningowy	Zbiór walidacyjny	Zbiór testowy
80%	10%	10%

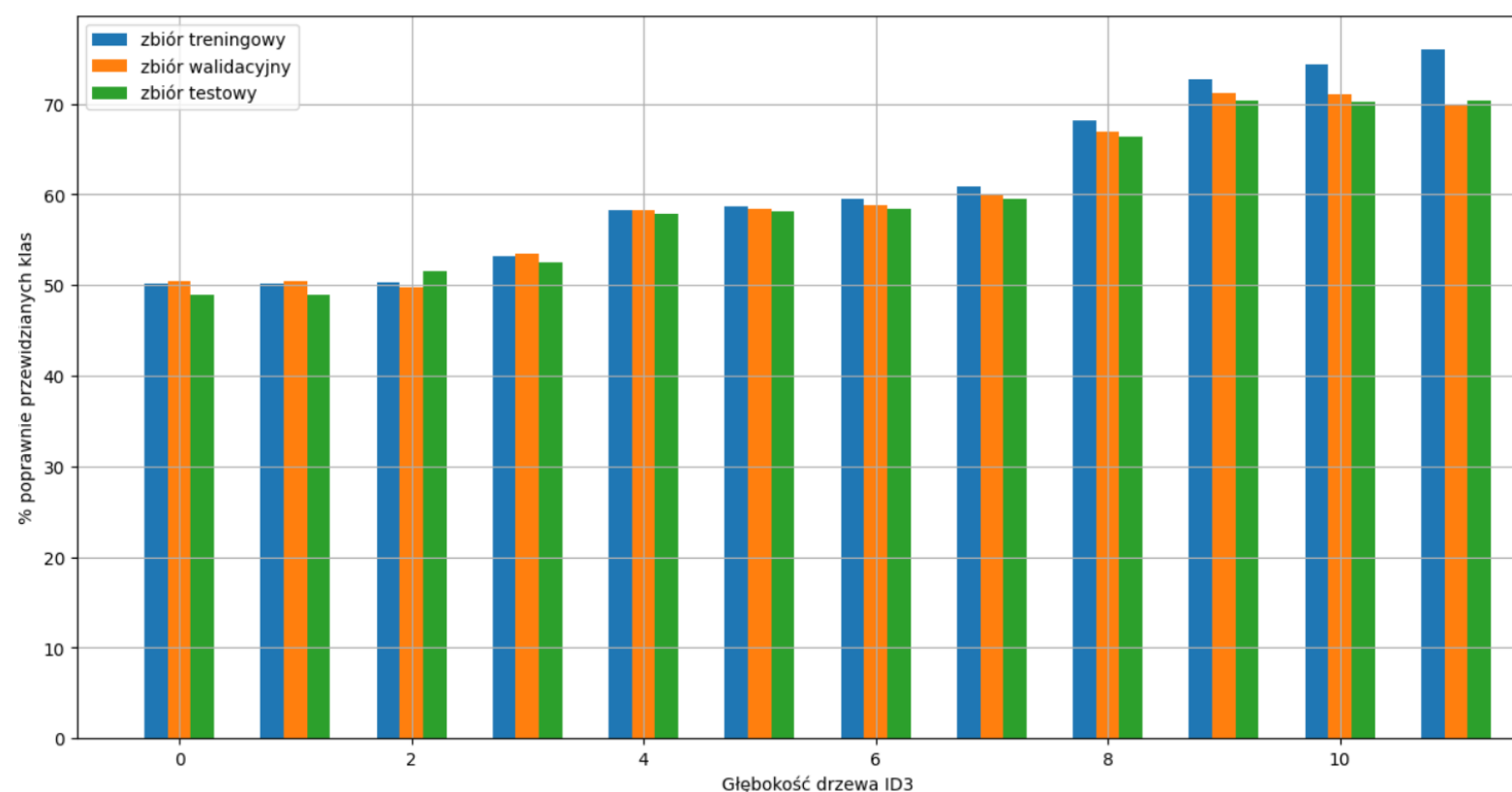
Tabela 2

Algorytm został zaimplementowany zgodnie z podanym pseudokodem oraz klasa z nim powiązana dziedziczy po abstrakcyjnym interfejsie Solver.

Oczekiwane wyniki

Spodziewam się, że wraz ze wzrostem głębokości budowanych drzew, algorytm będzie w stanie coraz lepiej przewidywać klasę dla X-ów. Należy jednak uważać na zjawisko przeuczenia.

Wyniki i komentarze



Wykres 1

Na wykresie nr 1 widać, że dla głębokości 0-2 algorytm jest całkowicie nieefektywny, bowiem jego skuteczność wynosi około 50%, czyli tyle co prawdopodobieństwo odgadnięcia wyniku, czy pacjent jest chory czy nie jest.

W przypadku głębokości 3 algorytm zaczyna wykazywać oznaki działania, gdyż jego % predykcji przewyższa 50% - wynosi około 53%.

Kolejną grupą rozwiązań są głębokości 4-7. Co ciekawe, działanie algorytmu jest dla nich zbliżone i kształtuje się na poziomie lekko poniżej 60% poprawnych predykcji.

Gwałtowny wzrost obserwujemy dopiero przy głębokości równej 8 – niemalże 10 punktów procentowych więcej niż w przypadku poprzedniej wartości hiperparametru.

Dla głębokości 9-11 skuteczność przekracza 70%. Warto też zwrócić uwagę, że dla największych głębokości zbiór walidacyjny i testowy przestają wykazywać większą skuteczność, natomiast predykcje na zbiorze treningowym stale rosną.

Na wykresie również widać, że zwykle zbiór walidacyjny osiągał lepsze wyniki od zbioru testowego. Oczywiście, algorytm najlepiej radził sobie na „sprawdzonych szlakach”, czyli zbiorze treningowym.

Przyjrzyjmy się także statystykom liczbowym z 15 obiegów algorytmu dla różnie generowanych zbiorów i najefektywniejszej głębokości (w przypadku zbioru treningowego) równej 11:

<i>Zbiór / Statystyka</i>	<i>Min</i>	<i>Średnia</i>	<i>Max</i>	<i>Odchylenie</i>
<i>Treningowy</i>	75.73	75.85	76.03	0.09
<i>Walidacyjny</i>	69.76	70.84	71.85	0.62
<i>Testowy</i>	69.59	70.67	71.22	0.42

Tabela 3

Jak widać w Tabeli nr 3, wartości dla zbioru treningowego są znacznie większe niż dla pozostałych, lecz wszystkie iteracje dają zbliżoną wartość – odchylenie jest znacznie mniejsze.

W przypadku zbiorów walidacyjnego i testowego średnie są zbliżone, lecz wartość maksymalna jest większa dla pierwszego z nich.

Przeanalizuję także statystyki liczbowe z 15 obiegów dla głębokości = 9:

<i>Zbiór / Statystyka</i>	<i>Min</i>	<i>Średnia</i>	<i>Max</i>	<i>Odchylenie</i>
<i>Treningowy</i>	72.38	72.64	72.77	0.1
<i>Walidacyjny</i>	70.14	70.99	72.15	0.64
<i>Testowy</i>	69.66	70.98	71.98	0.59

Tabela 4

Z tabeli nr 4 widać, że pomimo mniejszej efektywności dla zbioru treninowego, wszystkie statystyki liczbowe dla dwóch pozostałych zbiorów są lepsze niż dla maksymalnej głębokości. Jest to zatem głębokość graniczna, dla której nie zachodzi zjawisko przeuczenia się algorytmu.

Wnioski

1. W przypadku sprawdzania predykcji na zbiorze treningowym, algorytm najlepiej radzi sobie dla maksymalnej głębokości (w konkretnym przykładzie równej 11).
2. W przypadku predykcji dla zbiorów walidacyjnego i testowego głębokość=9 okazuje się być najlepsza.
3. Od głębokości 9 wartości predykcji dla zbiorów walidacyjnego i testowego pozostają na stałym poziomie pomimo jej wzrostu dla zbioru treningowego. Widać zjawisko przeuczania się algorytmu, podczas którego, nie generalizuje on wystraszająco wyników, lecz uczy się przewidywać cel dla konkretnych przypadków.
4. Głębokości od 0 do około 20% maksymalnej głębokości dla algorytmu są całkowicie nieefektywne i dają wynik wynikający równy prawdopodobieństwu jego odgadnięcia.
5. Największy wzrost efektywności algorytmu następuje między wartością głębokości 7 a 8.