

SPRAWOZDANIE

z ćwiczenia nr 3

„Regresja i klasyfikacja”

Przedmiot: Wprowadzenie do Sztucznej Inteligencji

Imię i nazwisko: Viktoriia Nowotka

W trakcie ćwiczenia realizowany algorytm ID3 dla drzew decyzyjnych z różną głębokością drzewa. Algorytm ID3 został zaimplementowany zgodnie z podanym pseudokodem. Warunki stopu są następujące: zwracamy liście drzewa, jeżeli

- napotkamy pusty zbiór danych wejściowych;
- wystąpi brak atrybutów do podziału;
- wszystkie etykiety są identyczne w zbiorze danych;
- osiągnięto maksymalną głębokość drzewa.

Podczas przewidywania w sytuacji, gdy w gałęzie drzewa nie istnieje potrzebnej wartości atrybutu zwracano etykieta, która najczęściej występuje w danym węźle.

Realizacja algorytmu była testowana na danych Cardio Vascular Disease Detection. Żeby dopasować kolumny wiek, wzrost, waga, ciśnienie do algorytmu ID3 były zastosowana dyskretyzacja poprzez zakresy:

- Wiek: 0-12, 13-25, 26-45, 45-60, 61+
- Ciśnienie skurczowe: 0-120, 121-140, 141+
- Ciśnienie rozkurczowe: 0-70, 71-90, 91+
- Wzrost: 0-150, 151-180, 181+
- Waga: 0-50, 51-80, 81+

Wyniki poprawności testów dla danych walidacyjnych raz predykcji danych testowych są następujące:

Tabela 1 – Dokładność walidacyjna drzewa decyzyjnego odnośnie głębokości

Głębokość drzewa	3	4	5	6	7	8	9	10	11	12	13
Dokładność walidacyjna	0.7263	0.7283	0.7303	0.7299	0.7278	0.7267	0.7225	0.7211	0.7193	0.7193	0.7193

Zgodnie z przedstawioną dokładnością algorytmu w tabeli nr 1 optymalna głębokość na zbiorze walidacyjnym jest 5.

Dokładność algorytmu ID3 dla danych testowych z głębokością 5 jest 0,7177, co jest bardzo zbliżona do dokładności walidacyjnej. To znaczy, że model nie jest zbyt przeuczony lub niedouczony.

Wnioski

W trakcie wykonania ćwiczenia nr 3 zaimplementowano algorytm ID3 dla drzew decyzyjnych. Stworzone rozwiązanie wykorzystano do kwalifikacji pacjentów na choroby kardiologiczne, co sprawdziło się dobrze.

Zostało zbadana głębokość drzewa na dokładność przewidywania klasy. Optymalną głębokością dla danego problemu jest 5. Mniejsza głębokość skutkowała niedouczeniem, a większa – przeuczeniem.

TESTY głębokości drzewa decyzyjnego z algorytmem ID3

