

Praca domowa 1

Termin oddania: 25.10.2023

1 Wstęp

Celem pracy domowej jest sprawdzenie jak poszczególne parametry modelu drzewa decyzyjnego wpływają na jego jakość predykcijną.

2 Zbiór danych

W tym celu posłużymy się zbiorem danych (`X.csv`, `y.csv`). Aby przygotować dane do dalszej pracy należy podzielić zbiór na treningowy i testowy w proporcji 7:3 ustawiając parametr `random_state = NUMERINDEKSU`. Następnie przy zadaniu z Sekcji 3.1 ponownie należy użyć parametru `random_state = NUMERINDEKSU` przy krosvalidacji.

3 Oczekiwany wynik

3.1 Eksperyment (6 punktów)

Przygotuj eksperyment ukazujący miarę dokładności drzewa decyzyjnego na zbiorze treningowym i testowym (wykorzystując krosvalidację, co najmniej 5-krotną na danych treningowych) w zależności od parametrów:

- kryterium podziału (`criterion`{ “*gini*”, “*entropy*”}),
- głębokość drzewa (`max_depth`),
- minimalna liczba obserwacji w liściu (`min_samples_leaf`).

Przejrzyj dokumentację dotyczącą budowy drzewa i spróbuj znaleźć inne parametry, które poprawią dokładność.

Przedstaw wnioski z eksperymentów w formie raportu (maksymalnie 2 strony A4).

3.2 Analiza jakości predykcyjnej modelu (4 punkty)

Na podstawie wyników z Sekcji 3.1 wybierz Twoim zdaniem najlepszy model, podaj uzasadnienie wyboru. Dla wybranego modelu na danych testowych wyznacz:

- macierz pomyłek,
- dokładność (ang. *accuracy*, *ACC*), czułość (ang. *sensitivity*, *recall*), precyzja (ang. *precision*),
- krzywą ROC, wartość AUC.

4 Szczegóły rozwiązania

Rozwiązanie powinno zawierać pliki:

- folder Kody zawierający wszystkie potrzebne kody do przygotowania rozwiązania zadania domowego,
- plik `NUMERINDEKSU_raport.pdf` opisujący przeprowadzony eksperyment oraz analizę wybranego modelu (maksymalnie 3 strony).

5 Ocena

Łączna liczba punktów do zdobycia jest równa 10, w tym:

- 3.1 Eksperyment (*6 punktów*)
 - jakość kodu (porządek, czytelność) - 1 punkt,
 - jakość eksperymentu - 3 punkty,
 - raport - 2 punkty.
- 3.2 Analiza jakości predykcyjnej modelu (*4 punkty*)
 - jakość kodu (porządek, czytelność) - 2 punkty,
 - wnioski - 1 punkt,
 - raport - 1 punkt.

6 Oddanie pracy domowej

Wszystkie punkty z sekcji *Szczegóły rozwiązania* należy umieścić w katalogu ZIP o nazwie NUMERINDEKSU_GR_PD1, gdzie

$$\text{GR} = \begin{cases} 1 & \text{dla środa, 12:15,} \\ 2 & \text{dla środa, 16:15.} \end{cases}$$

Tak przygotowany katalog należy przesłać na adres *anna.kozak@pw.edu.pl* do dnia 25.10.2023 do godziny 23:59. Tytuł wiadomości: *[WUM][PD1] Nazwisko Imię, Numer grupy: GR*.