Zadania - ewaluacja wyników

1. Dla zbioru danych IRIS wytrenuj model z użyciem:

A. metody cross\_validate dla 5 podzbiorów (cv), rozbij zadanie na 4 podprocesy (n\_jobs), w trybie komunikatywnym (verbose) używając modelu najbliższych sąsiadów z 3 sąsiadami,

B. wypisz wynik testu dla każdego z podzbiorów,

C. policz przeciętny wynik dla testu z użyciem metryki F1 score

D. wypisz wynik testu dla F1 score dla każdego z podzbiorów,

E. wypisz wynik testu dla czułości dla każdego z podzbiorów i jego przeciętną wartość,

F. użyj walidacji krzyżowej z 10 podzbiorami, czy wynik zmienił się?

G. zmierz czas treningu modelu.

2. Dla zbioru danych z IRIS wykonaj trening modelu z użyciem metody GridSearch do odnalezienia jego najlepszej konfiguracji:

A. przygotuj model z użyciem algorytmu lasów losowych random forest i sprawdź dla której z wszystkich kombinacji parametrów uzyska on największą precyzję klasyfikacji. Sprawdź go dla wartości: max\_depth z zakresu: 1, 2, 4, 8, ..., 25, 64 i n\_estimators z zakresu 1,2,3,

B. policz przeciętny wynik dla testu z użyciem metryki F1 score,

C. wypisz wynik testu dla F1 score dla każdego z podzbiorów,

D. wypisz wynik testu dla precyzji dla każdego z podzbiorów i jego przeciętną wartość,

E. wypisz najlepszą uzyskaną konfigurację dla uzyskanych parametrów

F. zmodyfikuj zadanie i użyj walidacji krzyżowej z 10 podzbiorami, czy wynik zmienił się?

G. zmierz czas treningu dla tego zadania.

3. Dla zbioru danych diabetes wytrenuj model używając:

A. metody cross\_validate dla 7 podzbiorów (cv), rozbij zadanie na 2 podprocesy (n\_jobs), w trybie komunikatywnym (verbose) używając

modelu najbliższych sąsiadów z 5 sąsiadami,

B. wypisz wynik testu dla każdego z podzbiorów,

C. policz przeciętny wynik dla testu z użyciem metryki R2

D. wypisz wynik testu dla R2 dla każdego z podzbiorów,

E. wypisz wynik testu dla metryki MAE dla każdego z podzbiorów i jego przeciętną wartość,

F. użyj walidacji krzyżowej z 10 podzbiorami, czy wynik zmienił się?

G. zmierz czas treningu modelu.

4. Dla zbioru danych diabetes wykonaj model z użyciem:

A. użyj metody Randomized Search do odnalezienia najlepszej konfiguracji modelu:

przygotuj model z użyciem algorytmu lasów losowych i sprawdź dla której ze 100 wybranych losowo kombinacji parametrów uzyska on największą precyzję klasyfikacji. Sprawdź go dla wartości: max\_depth z zakresu: 1, 2, 4, 8, 9, 12, 81, 1024 i n\_estimators z zakresu 3, 6,

B. wypisz najlepszą uzyskaną konfigurację dla uzyskanych parametrów

C. policz przeciętny wynik dla testu z użyciem metryki R2

D. wypisz wynik testu dla R2 dla każdego z podzbiorów,

E. wypisz wynik testu dla metryki MAE dla każdego z podzbiorów i jego przeciętną wartość,

F. zmodyfikuj zadanie i użyj walidacji krzyżowej z 10 podzbiorami, czy wynik zmienił się?

G. zmierz czas treningu dla treningu wykonanego dla 200 testowanych kombinacji parametrów metody.

W kolejnych zadaniach zmienia się tylko zbiór danych (... i wyniki):

5. Wykonaj polecenia zad.1 dla zbioru danych z zadania 9 z listy 3 (klasyf.: 5 kol./500 w.)

6. Wykonaj polecenia zad.2 dla zbioru danych z zadania 9 z listy 3

7. Wykonaj polecenia zad.3 dla zbioru danych z zadania 9 z listy 3

8. Wykonaj polecenia zad.4 dla zbioru danych z zadania 10 z listy 3 (regr.: 4 kol./400 w.)

9. Wykonaj polecenia zad.5 dla zbioru danych z zadania 10 z listy 3

10. Wykonaj polecenia zad.6 dla zbioru danych z zadania 10 z listy 3

W zadaniach warto skorzystać metod z modułu: sklearn.model\_selection - https://scikit-learn.org/stable/modules/classes.html#module- sklearn.model selection