

Code review for deepRession

Szymon Gut
Wiktor Jakubowski
Maria Kedzierska
Maciej Orłowski

April 2022

1 Uwagi ogólne

1. Dla funkcji nie działa opcja `?nazwa_funkcji`.
2. Niekonsekwencja w Enterach między komentarzem a nazwą funkcji.
3. Dobrze byłoby się zdecydować na trzymanie napisów w ten sam sposób w `'` albo `"`.
4. Część komentarzy jest po polsku, warto je zmienić na angielski.
5. Odkomentować lub usunąć zakomentowane linijki kodu.
6. W przykładach jest brana zła funkcja `rmse`, przez co rzuca błędem.

2 Uwagi do poszczególnych funkcji

2.1 `to_DataFrame()`

1. Nazwa funkcji powinna być utrzymana w jednej konwencji, tzn. `camelCase` albo z podłogami. Nasza propozycja to zmiana nazwy funkcji na `to_data_frame`.

2.2 `null_exterminator()`

1. Literówka w komentarzu w słowie `'target'`.

2.3 `scale_continuous()`

1. Nie sprawdza, czy któraś z kolumn nie jest numeryczna.
2. Warto dodać przykłady, gdzie `to_scale` zawiera jakąś wartość NA lub kolumnę kategoriową.
3. W przypadku wykrycia wartości NA printowany komentarz jest nieadekwatny do rzeczywistej sytuacji.

2.4 `train_test_split()`

1. Źle inicjowane `smp_size`, bez wykorzystania parametru `train_size`.
2. Powinna być spacja po '=' w definicji parametrów funkcji.

2.5 `preprocessing()`

1. Brak spacji po przecinkach w `one_hot_encode`.

3 Podsumowanie

1. Kod realizuje cel jakim jest przygotowanie danych do wytrenowania przez model `RandomForest`.
2. Kod spełnia wymagania omówione podczas prezentacji.
3. Kod generalnie jest zgodny z wytycznymi stylistycznymi poza drobnymi wyjątkami np. brak spacji po przecinkach lub = w niektórych miejscach.
4. Funkcja `one_hot_encode` jest dość mało czytelna, można byłoby dodać bardziej opisowe nazwy zmiennych, spróbować jakoś zmniejszyć liczbę warunków logicznych lub dodać komentarze.
5. Komentarze i dokumentacja są bardzo przejrzyste. W dokumentacji brakuje czasem przykładów, które testowałyby specyficzne przypadki użycia funkcji.
6. W kodzie zdarzyły się dwa błędy logiczne.
7. Udało się odtworzyć wszystkie przykłady poza ostatnim, gdzie jest wyliczane `rmse`.
8. Funkcja `train_test_split` nie zadziałała na naszych danych.