

Praca domowa 3 - code review grupy WTF

Bujakowski Hubert

May 2022

1 Czy ten kod osiąga cel, który postawiono?

Tak, wszystko co zostało zawarte na GitHub'ie w KM2 wydają się osiągać postawiony cel.

2 Czy w kodzie są jakieś oczywiste błędy logiczne?

Nie, uważam że oczywiste błędy od razu rzuciłyby się w oczy, tutaj tak nie było.

3 Czy patrząc na wymagania zawarte podczas prezentacji są one w pełni zaimplementowane?

Nie jestem w stanie odpowiedzieć na postawione pytanie, na GitHub'ie jest tylko kod, bez prezentacji – przez co ciężko jest określić czy wszystkie wymagania/założenia zostały spełnione.

4 Czy są jakieś obszary, w których kod mógłby zostać poprawiony? (skrócić, przyspieszyć, itp.)

- Tak, czasami przeklejanie kodu skutkuje błędami, np. w pliku **run.ipynb** drugi warunek nie powinien znajdować się w pętli. Tutaj nie jest to odczuwalne ale dla większej ilości kolumn już tak.

```
def pipeline(X,y):  
    for feature in X.columns:  
        if X[feature].dtype == 'object':  
            X[feature] = X[feature].astype('category')  
        if y.dtype == 'object':  
            y = y.astype('category')  
    return X, y
```

- Powtarzanie kodu, moim zdaniem lepszym pomysłem byłoby wstawić funkcje do np. pewnego pliku .py, który byłby importowany na początku danego notebook'a. Wtedy m.in. nie doszłoby do pomyłki

5 Styl kodu

W pliku **exec.py** napotkałem kilka przykładów które nie są zgodne z istniejącymi wymaganiami stylistycznymi (PEP-8), m.in:

- brak 2 pustych linii (tylko 1) po wszystkich importach

```
...
import warnings

def funkcja(X,y, fold_time = 300, time_per_model = 75, n_folds = 10):
    ...
```

- 9 linia - brak whitespace (jw.)
- 17,18 linia kodu – za dużo pustych elementów w środku funkcji
- 20 linia – wcięcie nie jest wielokrotnością 4
- 24 linia – blankspace
- 36 linia – brak whitespace
- 37 linia jest za długa (więcej niż 79 znaków)

6 Czy dokumentacja i komentarze są wystarczające?

W niektórych plikach jest 0 komentarzy, w niektórych komórki, które prowadzą czytelnika co w danym pliku zostało zrobione. Ogólne odczucia są pozytywne, kod jest na tyle przejrzysty że da się go zinterpretować bez większego zagłębienia się w dany temat.

7 Czy udało się odtworzyć zamieszczone przykłady w kodzie?

Wszystkich przykładów nie testowałem ale dla przykładu na datasetcie 'kr-vs-kp' otrzymałem odp. wyniki dla 2 foldów: 0.99984, 0.99812 (wg metryki roc_auc)

8 Czy udało się użyć przygotowanych kodów na nowych danych?

Tak, przetestowałem framework na zbiorze 'Indian Liver Patient', roc_auc wyniosło: 0.720284

9 Podsumowanie

- Kod osiąga postawiony cel
- Brak poważnych błędów
- Kod nie jest zgodny z PEP-8
- Brak dokumentacji
- Kod da się wykorzystać przy pracy na innych zbiorach