

Code review - AutoGluon

Łukasz Tomaszewski

4 maja 2022

1 Czy ten kod osiąga cel, który postawiono?

1.1 Opis

Funkcja miała przyjmować zbiór danych objaśniających i kolumnę objaśnianą. Oceniany kod pozwala natomiast jedynie na wczytanie danych za pomocą adresu URL.

Wczytane dane poddane są preprocessingowi polegającemu na decodowaniu kolumn typu **object** na **utf-8**.

Następnie tworzone są foldy, które przekazywane są do predyktora. Zwracane są najlepsze wyniki dla każdego z foldów.

1.2 Wnioski

Wystarczy zmodyfikować funkcję tak, aby przyjmowała ona zmienne objaśniające i kolumnę objaśnianą.

2 Czy w kodzie są jakieś oczywiste błędy logiczne?

Jedyne błędy logiczne polegają na impotowaniu pakietów. Pakiet numpy nigdy nie jest używany. Importowany jest także cały pakiet autogluon (wczytany jako ag nigdy nie jest użyty), a następnie importowana jest tylko jego część.

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

Brakuje tylko możliwości uruchomienia funkcji korzystając ze zmiennych objaśniających i zmiennej objaśnianej.

4 Czy kod jest zgodny z istniejącymi wytycznymi stylistycznymi?

W większości kod jest zgodny z PEP 8. Zmienne oraz funkcje mają logiczne nazwy. Występują pojedyncze błędy polegające na braku spacji lub linii przerwy.

Znaleziono błędy:

- PEP 8: E302 expected 2 blank lines, found 1 :11
- PEP 8: E261 at least two spaces before inline comment :34
- PEP 8: E131 continuation line unaligned for hanging indent :35
- Variable in function should be lowercase :56
- PEP 8: E261 at least two spaces before inline comment :62

5 Czy są jakieś obszary, w których kod mógłby zostać poprawiony?

Należałoby pominąć importowanie pakietów, które nie są wykorzystywane w kodzie. Uważam, że lepiej byłoby gdyby numer folda printowany byłby przed szukaniem modelu.

6 Czy dokumentacja i komentarze są wystarczające?

Wszystko jest zrozumiałe. Uważam, że warto umieścić możliwe wartości zmiennej preset w dokumentacji, zamiast w komentarzu. Poza tym funkcja jest dobrze opisana. Wiadomo co robią poszczególne fragmenty.

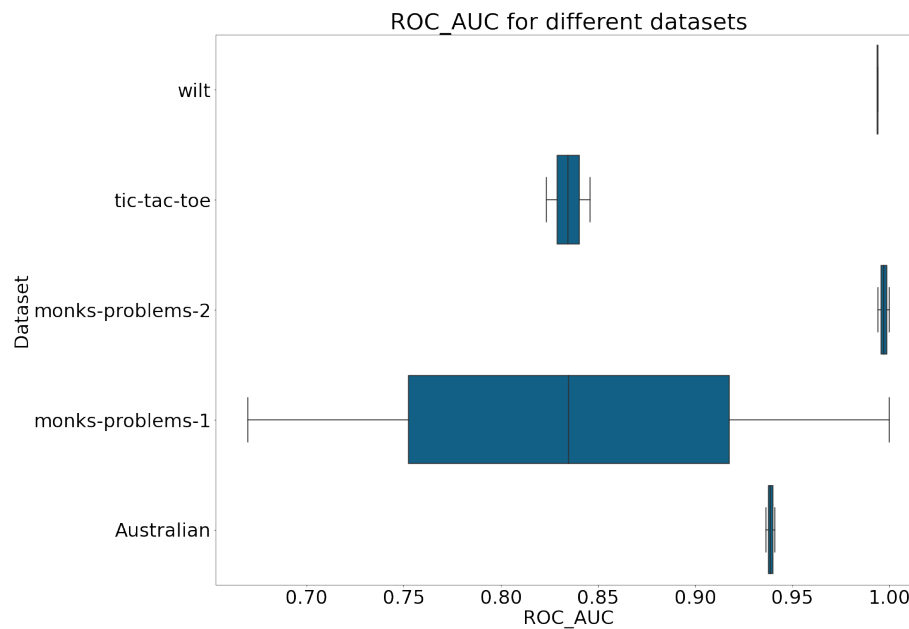
7 Czy udało się odtworzyć zamieszczone przykłady w kodzie?

Udało się uruchomić funkcję dla wszystkich binarnych tasków z OpenML.

8 Czy udało się użyć przygotowanych kodów na nowych danych?

8.1 Opis

Jako, że kod oferuje tylko uruchomienie zadań przy pomocy adresu URL, to funkcja została uruchomiona dla pięciu tasków z OpenML o ID równym: 9914, 145804, 146065, 146064, 125923. Dla każdego zbioru zostały policzone 2 foldy.



Rysunek 1: Wyniki dla tasków z OpenML (2 foldy)

8.2 Wnioski

Funkcja zadziałała na wybranych taskach. Dla czterech z nich zwróciła zbliżone wyniki dla obydwu foldów. Dla jednego tasku wyniki znacznie się różniły.

9 Podsumowanie

Funkcja wymaga przebudowy, aby być w stanie obsługiwać zmienne objaśniające i zmienną objaśnianą. Poza tym wymaga ona drobnych popraw: usunięcie importowania niepotrzebnych pakietów oraz drobne zmiany wizualne, aby spełniać standardy PEP8. Poza tym funkcja działa bardzo dobrze, zarówno na przygotowanych zbiorach, jak i na nowych zbiorach pobranych z OpenML.