

Второе ДЗ

Машинное обучение в продакшене

Всем привет, с праздниками!

Настало время добавить еще одну домашку, получить за нее можно **15 баллов (+ бонусные)**.

СРОКИ(после soft -- баллы * 0.6):

soft deadline -- **до 16 мая включительно**

hard deadline -- **до 23 мая включительно**

Механика проверки такая же(добавляем преподавателя и студента, за сделанное ревью 2 балла).

Задание:

В прошлом ДЗ вы обучили модель для решения задачи классификации(по умолчанию использовался датасет <https://www.kaggle.com/ronitf/heart-disease-uci>), новое задание, это обернуть ее в вид, пригодный для использования в режиме онлайн

0) ветку назовите homework2, положите код в папку online_inference

1) Оберните inference вашей модели в rest сервис(вы можете использовать как FastAPI, так и flask, другие желательно не использовать, дабы не плодить излишнего разнообразия для проверяющих), должен быть endpoint /predict (3 балла)

2) Напишите тест для /predict (3 балла)

(<https://fastapi.tiangolo.com/tutorial/testing/>, <https://flask.palletsprojects.com/en/1.1.x/testing/>)

3) Напишите скрипт, который будет делать запросы к вашему сервису -- 2 балла

4) Сделайте **валидацию** входных данных (например, порядок колонок не совпадает с трейном, типы не те и пр, в рамках вашей фантазии) (вы можете сохранить вместе с моделью доп информацию, о структуре входных данных, если это нужно) -- 3 **доп** балла

<https://fastapi.tiangolo.com/tutorial/handling-errors/> -- возвращайте 400, в случае, если валидация не пройдена

5) Напишите dockerfile, соберите на его основе образ и запустите локально контейнер(docker build, docker run), внутри контейнера должен запускать сервис, написанный в предыдущем пункте, закоммитьте его, напишите в readme корректную команду сборки (4 балл)

6) Оптимизируйте размер docker image (3 **доп** балла) (опишите в readme.md что вы предприняли для сокращения размера и каких результатов удалось добиться) -- https://docs.docker.com/develop/develop-images/dockerfile_best-practices/

7) опубликуйте образ в <https://hub.docker.com/>, используя **docker push** (вам потребуется зарегистрироваться) (2 балла)

8) напишите в readme корректные команды docker pull/run, которые должны привести к тому, что локально поднимется на inference ваша модель (1 балл)

Убедитесь, что вы можете протыкать его скриптом из пункта 3

5) проведите самооценку -- 1 **доп** балл

Примеры с пары:

- https://github.com/made-ml-in-prod-2021/inference_examples

Полагаю, по мере выполнения этого ДЗ будет просматриваться лекция по докеру и рест -- **заполняйте опрос** <https://forms.gle/DeriST61D1AYmtNR7>

Норм?
норм
сложно
слишком легко
Воздержаться