Лабораторная работа №1

КЛАССИЧЕСКИЙ ЖИЗНЕННЫЙ ЦИКЛ РАЗРАБОТКИ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ

В качестве источника данных использовался 11 вариант. Решалась задача классификации текста. Для этого использовался TF-IDF + логическая регрессия.

Ссылка на репозиторий git: <https://github.com/Saivaks/big_data_lab_1>

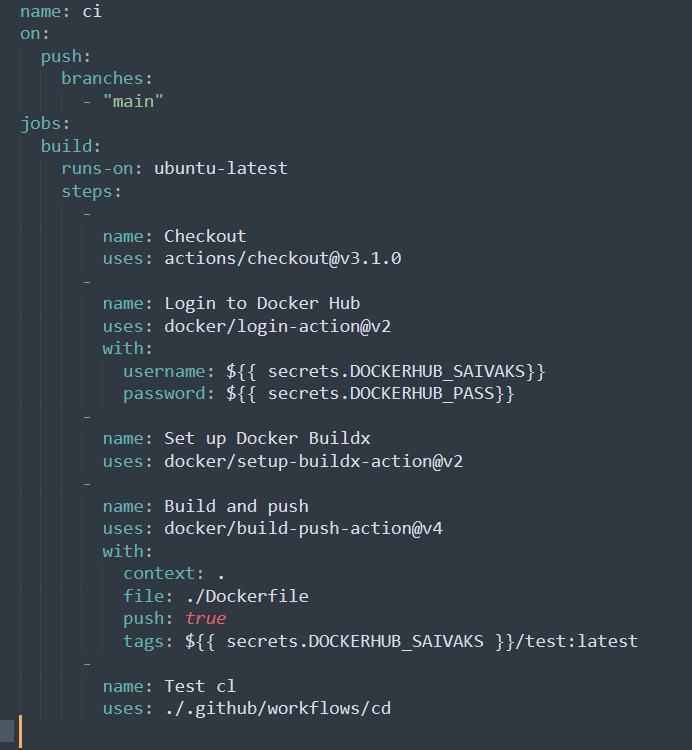
Ссылка на докера хаб: <https://hub.docker.com/repository/docker/saivaks/test/general>

Для тестирования использовался фрейморк pytest. Проверялись данные на соответствие: длины входного и выходного датасета, его типов, количества категорий.

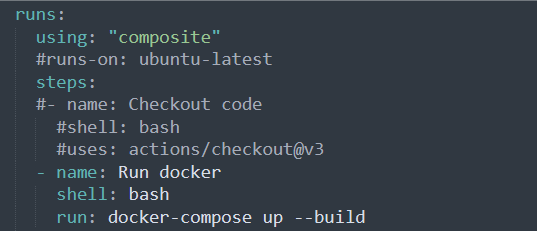
Был создан Dockerfile и docker-compose.yml которые содержат модель, а также в нем производится ее запуск с необходимыми зависимостями и параметрами.

Был создан CI который отправляет на докер хаб образ докера после каждого push и CL который проводит запуск докера и тестирование.

Код CI:



Код CL



Результаты отработки пайплана (в том числе и тестирования):

