Инструкция.

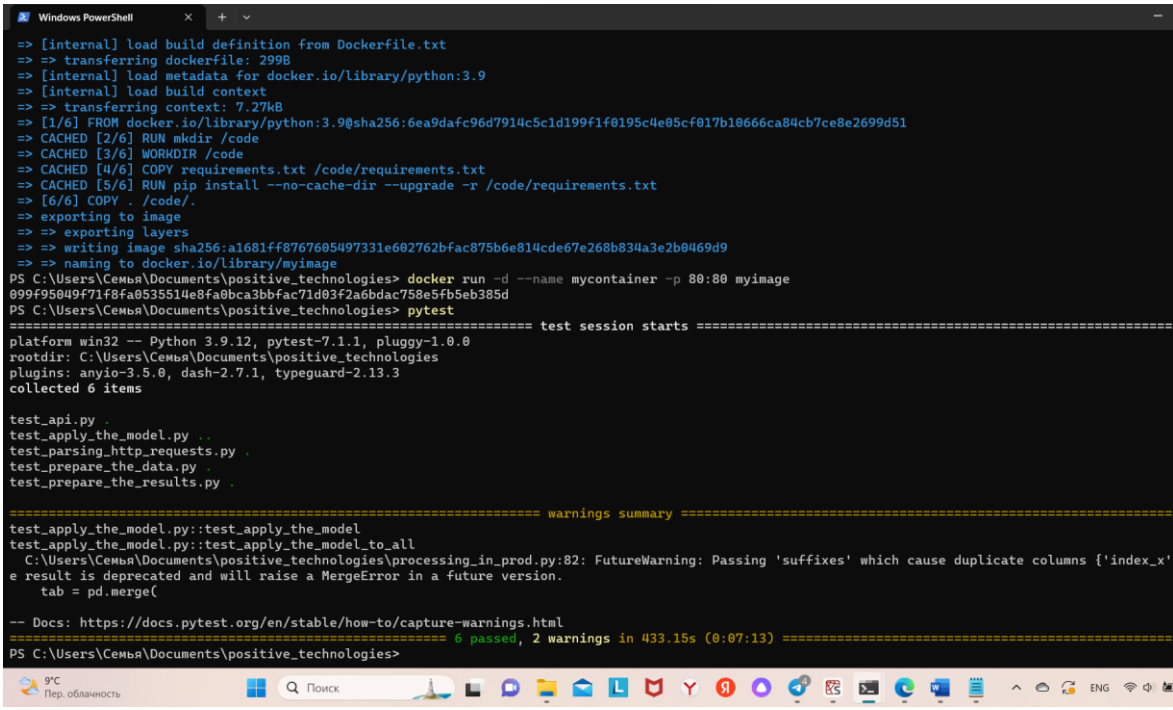
Список файлов.

- **anomaly_mapping.csv**, один из служебных csv-файлов, входящих в состав модели
- **api.py**, файл, содержащий код основной функции **separate_the_flees**
- **app.py**, файл, содержащий код эндпоинта **predict**
- Dockerfile.txt, докерфайл
- **eda_analysis.ipynb**, ноутбук, в котором строится модель и поясняется её выбор
- **hdbscan_model.zip**, сама модель, которую надо вручную достать из архива
- **model.py**, один из служебных файлов
- **part_10.csv**, исходные данные
- **popularity_stats_CLIENT_USERAGENT**, один из служебных csv-файлов, входящих в состав модели
- **popularity_stats_EVENT_ID**, один из служебных csv-файлов, входящих в состав модели
- **popularity_stats_MATCHED_VARIABLE_NAME**, один из служебных csv-файлов, входящих в состав модели
- **popularity_stats_MATCHED_VARIABLE_SRC**, один из служебных csv-файлов, входящих в состав модели
- **popularity_stats_MATCHED_VARIABLE_VALUE**, один из служебных csv-файлов, входящих в состав модели
- **processing.py**, один из служебных файлов
- **processing_for_rest.py**, один из служебных файлов
- **processing_in_prod.py**, один из служебных файлов
- **requirements.txt**, зависимости
- **test_api.py**, тест на api, оттуда можно взять пример вызова по REST
- **test_apply_the_model.py**, два теста на модель, один из них на всём датасете
- **test_parsing_http_requests.py**, один из тестов
- **test_prepare_the_data.py**, один из тестов
- **test_prepare_the_results.py**, один из тестов

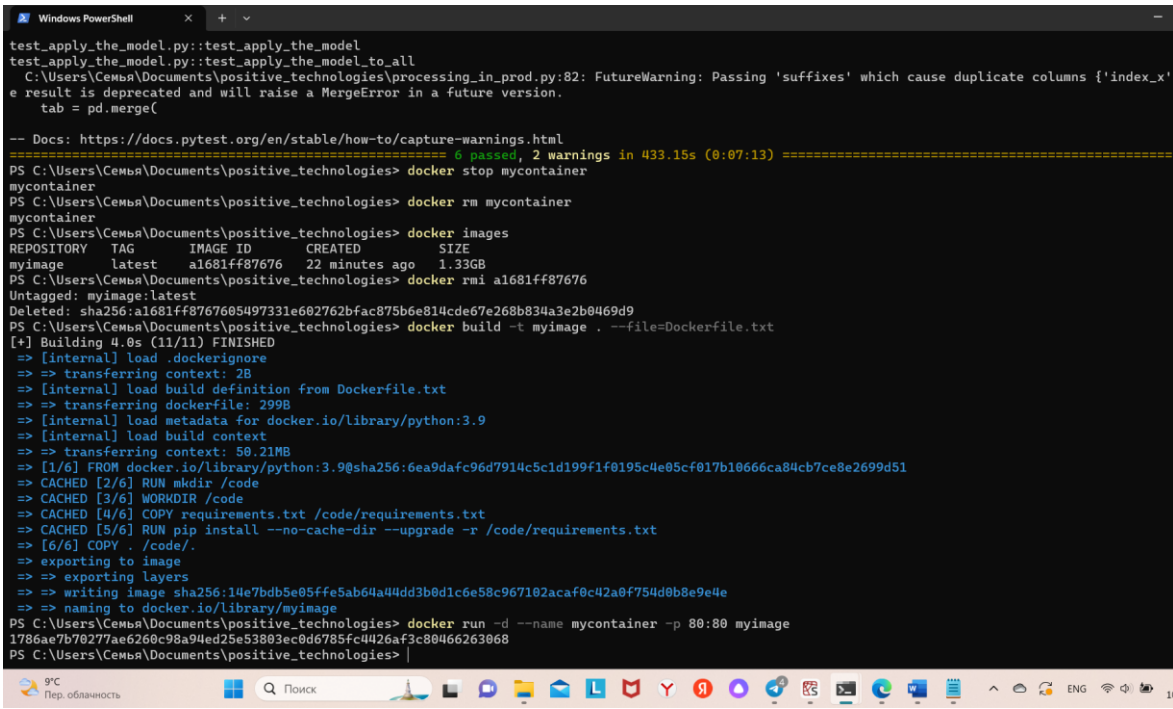
- **анализ.txt**, файл который для каждого аномального кластера содержит статистику этого кластера и один из примеров (нужен для интерпретируемости результатов)

Тестирование.

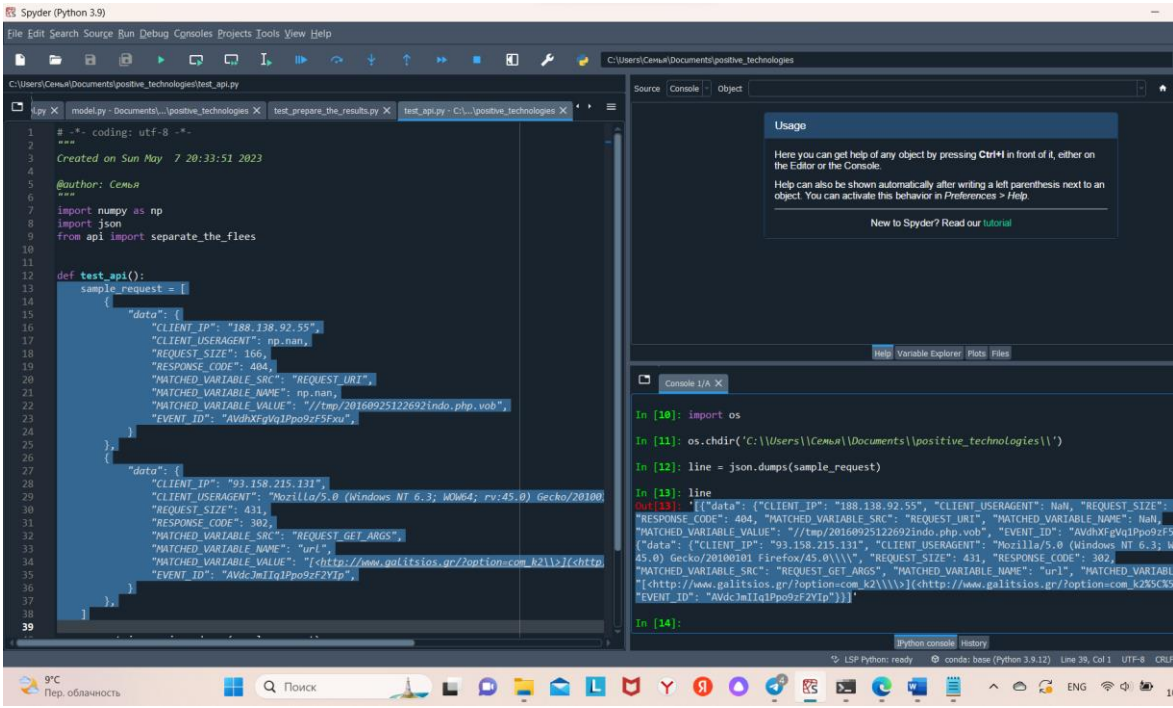
Тесты запускаются через **pytest**:

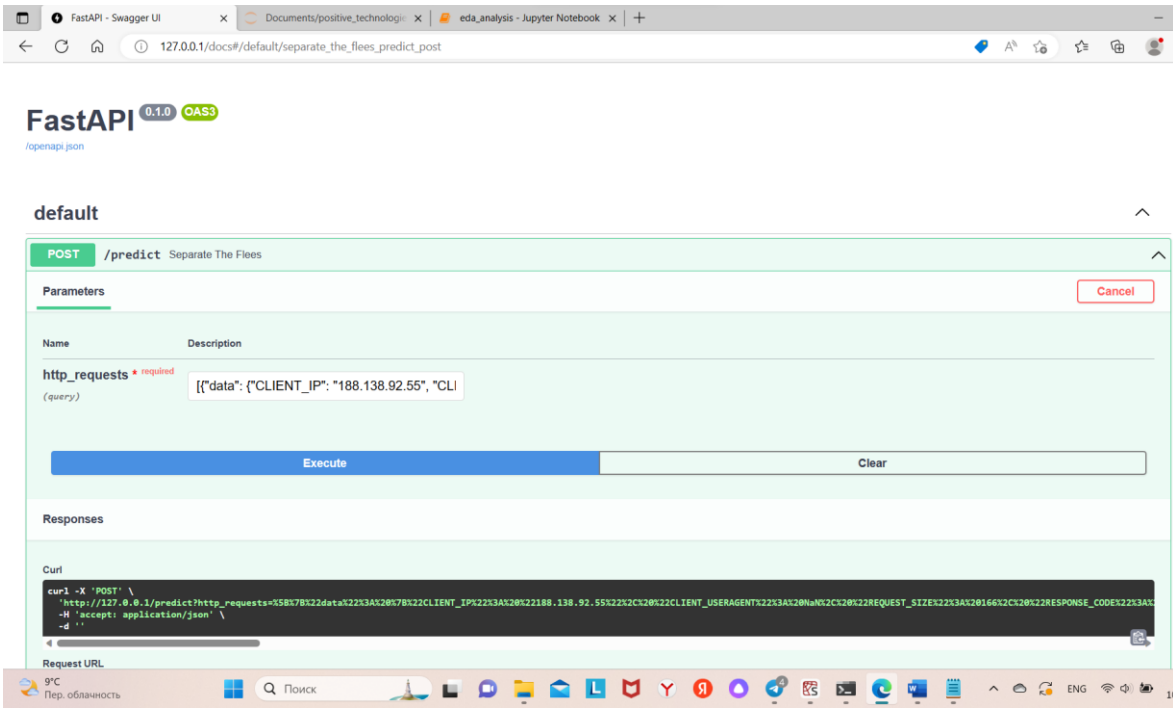


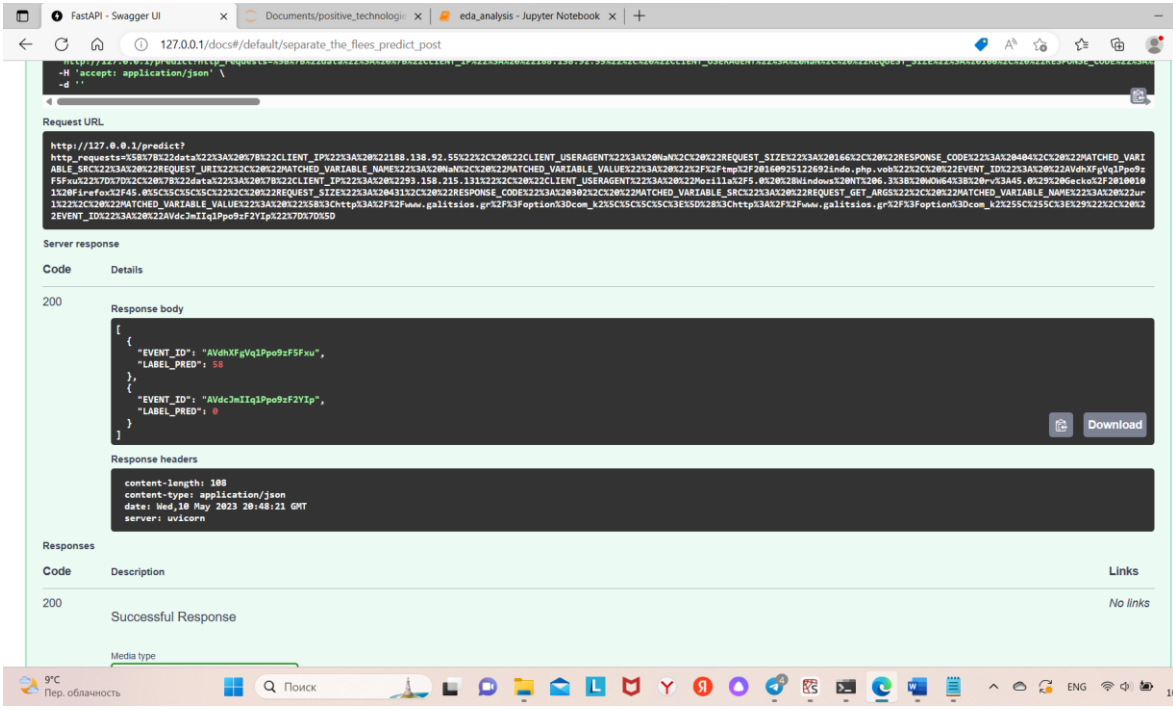
Запуск через докер.



Идём в файл **test_api.py**, берём пример **sample_request**. Сериализуем его с помощью **json.dumps**.


Spyder (Python 3.9)
File Edit Search Source Run Debug Consoles Projects Tools View Help
Usage
Here you can get help of any object by pressing Ctrl+I in front of it, either on the Editor or the Console.
Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.
New to Spyder? Read our tutorial
Console 1/A
In [10]: import os
In [11]: os.chdir('C:\\Users\\Семья\\Documents\\positive_technologies\\')
In [12]: line = json.dumps(sample_request)
In [13]: line
In [14]:


FastAPI - Swagger UI
127.0.0.1/docs#/default/separate_the_flees_predict_post
FastAPI
0.1.0
OAS3
/openapi.json
default
POST
/predict
Separate The Flees
Parameters
Cancel
Name
Description
http_requests
required
(query)
Execute
Clear
Responses
Curl
Request URL



Интерпретация выходных параметров.

Если в **LABEL_PRED** стоит пропущенное значение, то это не аномалия.

Если в **LABEL_PRED** стоит число от 0 до 53, то это один из аномальных кластеров, полученных HDBSCAN.

Если в **LABEL_PRED** стоит число от 54 до 58, то это один из аномальных кластеров, соответствующих пропущенным значениям.

Пример как автоматически присылать ответы.

Файл **test_apply_the_model.py**, функция **test_apply_the_model_to_all** показывает как запустить основную логику автоматически на csv-файле.