

Описание проекта «ai24_ecg»

Структура проекта

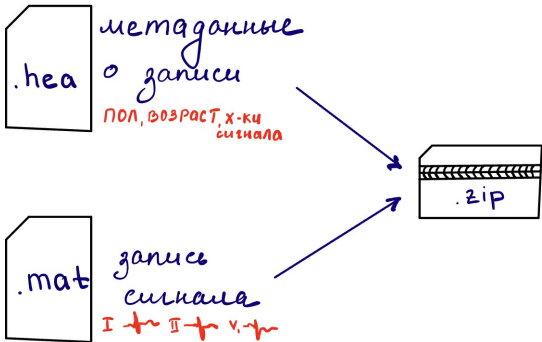
Проект разделен на backend - реализован посредством FastAPI , frontend - фреймворк Streamlit , а также организовано логирование посредством logstash

Репозиторий организован следующим образом:

```
.
├── .gitignore  исключения определённых файлов из системы контроля версий Git
├── ANOTHER    дополнительные файлы/модули проекта с предыдущих чекпоинтов
├── LICENSE    тип лицензии проекта
├── README.md  файл с описанием проекта
├── docker-compose.yml  файл конфигурации docker-compose
├── backend    директория с серверной частью приложения
│   ├── main.py    основной исполняемый файл серверной части
│   ├── models.py  модели данных
│   ├── training.py скрипт для обучения модели
│   ├── logger_config.py  настройки логирования
│   ├── requirements.txt  список зависимостей для серверной части
│   ├── utils.py   вспомогательные функции
│   ├── snomed-ct.csv  файл данных с декодировкой целевой переменной
│   └── Dockerfile  Docker-конфигурация для серверного приложения
├── frontend   директория с клиентской частью приложения на Streamlit
│   ├── app.py  основной исполняемый файл клиентской части
│   ├── logger_config.py  настройки логирования для фронтенда
│   ├── requirements.txt  список зависимостей для клиентской части
│   └── Dockerfile  Docker-конфигурация для клиентского приложения
├── logstash    директория с Logstash для обработки и анализа логов
├── logstash.conf  конфигурация для Logstash
└── Dockerfile  Docker-конфигурация для Logstash
```

Характеристика используемых данных

Для корректной работы с приложением требуется предоставление .zip-архива с вложенными данными двух расширений: .hea и .mat. Подробное



назначение файлов приведено на схеме.

ВНИМАНИЕ! Образцы данных для использования приложения приведены в этом блоке:

- [полный набор данных для EDA и обучения моделей Georgia.zip](#)
- [комплект тестовых данных для инференса test_ecg.zip](#)

Описание backend части

API реализовано с использованием фреймворка FastAPI. Основная цель — предоставление функциональности для обучения моделей, работы с данными и предоставления результатов.

Подключаемые модули:

- FastAPI : Фреймворк для создания API
- pydantic : Для валидации данных
- utils : Пользовательские утилиты для обработки файлов и данных
- training : Модуль для управления процессом обучения моделей
- logger_config : Настройки логирования

Конечные точки

1. GET /status

Описание: Проверка доступности API. Ответ: JSON с сообщением о том, что сервер работает.

2. POST /upload

Описание: Загрузка файла на сервер для обработки.

Параметры:

- `file`: Загрузить файл через HTTP-запрос.

Ответ: Подтверждение успешной загрузки файла.

3. POST /train

Описание: Запуск процесса обучения модели.

Параметры:

- Настройки гиперпараметров обучения (передаются в теле запроса).

Ответ: JSON-объект с ID эксперимента или сообщением о старте обучения.

4. GET /experiments

Описание: Получение списка всех экспериментов.

Ответ: Список ID экспериментов и их описания.

5. GET /metrics/{experiment_id}

Описание: Получение метрик для конкретного эксперимента.

Параметры:

- `experiment_id`: Идентификатор эксперимента.

Ответ: Метрики в формате JSON.

6. GET /curves/{experiment_id}

Описание: Получение данных обучения для заданного эксперимента

Параметры:

- `experiment_id`: Идентификатор эксперимента.

Ответ: Данные для построения графиков.

7. GET /eda

Описание: Проведение анализа данных (EDA).

Ответ: JSON с результатами анализа.

Особенности:

1. **Логирование**
 - Логи записываются в файлы и сохраняются в папке `backend/logs`.- Используется модуль `logger_config` для ротации логов.
2. **Обработка ошибок**
 - Все исключения возвращаются в формате JSON с соответствующим HTTP-статусом.- Пример: Если загружается файл неправильного формата, возвращается ошибка 400.
3. **Работа с файлами**
 - Загрузка файлов ограничена определёнными типами, указанными в `utils.ALLOWED_EXTENSIONS`.- Загрузка и хранение файлов осуществляется в отдельной папке.

Описание frontend части

Приложение `Streamlit` реализовано для взаимодействия пользователя с системой обработки сигналов ЭКГ и анализа состояния здоровья.

Основные компоненты приложения

1. **Заголовок** Отображается главный заголовок приложения: "Приложение для анализа сигналов ЭКГ и определения состояния здоровья человека".
2. **Работа с состоянием** Используется механизм `st.session_state` для сохранения состояния приложения (например, имени загруженного файла).
3. **Загрузка данных** Пользователь может загрузить файл для анализа (например, CSV с данными ЭКГ). Файл сохраняется в `st.session_state["dataset_name"]`.
4. **Визуализация данных**
 - Графики и визуализация:
 - Используется библиотека `Plotly` для построения интерактивных графиков.

- Возможна визуализация статистики и сигналов ЭКГ.
- Дополнительные библиотеки:
- `matplotlib` и `seaborn` для построения плотностей, корреляций и других графиков.

5. Взаимодействие с API

- Взаимодействие с серверной частью осуществляется через HTTP-запросы к `BACKEND_URL`.
- Примеры функциональности:
- Отправка данных на сервер для предсказания.
- Запрос метрик или результатов экспериментов.

6. Ключевые функции приложения

- **Запуск предсказания:** Кнопка, которая отправляет данные на сервер через API и выводит результат.
- **Отображение результатов:** Графическая визуализация данных, полученных от API.
- **Логирование:** Логгируется каждое действие пользователя с помощью `logger`.

7. Интерфейс пользователя

- Удобные элементы управления:
- Кнопки (`st.button`), загрузка файлов (`st.file_uploader`), слайдеры и текстовые поля.
- Динамическое обновление интерфейса в зависимости от действий пользователя.

Особенности

1. Гибкость визуализации

- Поддержка различных форматов данных (включая CSV).
- Построение нескольких типов графиков для анализа данных.

2. Простота использования

- Пользовательский интерфейс создан для интерактивной работы даже без технической подготовки.
- Подробные уведомления и сообщения при взаимодействии с элементами.

3. Логирование

- Используется модуль `logger_config` для фиксации событий в приложении.
- Логи сохраняются на сервере.

Инструкция по использованию

Инструкция для запуска проекта

1. Убедитесь, что на вашей системе установлен Docker и Docker Compose.
2. Склонируйте проект:

```
git clone https://github.com/akiltrebreg/ai24_ecg.git
cd ai24_ecg-streamlit_patch
```

3. Запустите проект:

```
docker-compose up -d
```

4. Откройте приложение: Интерфейс Streamlit доступен по адресу: <http://localhost:8501>, также Вы можете воспользоваться развернутым нами сервисом по адресу: <http://195.133.13.244:8501/>

Инструкция для работы с приложением Streamlit

Следуйте этим шагам для использования приложения:

1. **Запуск приложения** Откройте ваш браузер и перейдите по адресу: <http://localhost:8501>
2. **Главный экран** Вы увидите заголовок: "Приложение для анализа сигналов ЭКГ и определения состояния здоровья человека". Под ним отображается краткое описание приложения.
3. **Загрузка данных** В разделе "Загрузите ZIP-файл для обучения" выполните следующие действия:
 - Нажмите кнопку "Browse files".
 - Выберите файл, содержащий данные ЭКГ.
 - Подождите, пока файл загрузится и появится надпись "Файл успешно загружен". Название файла будет отображено в интерфейсе.
 - Под ним отображается краткое описание файла

ВНИМАНИЕ! Образцы данных для использования приложения приведены в этом блоке:

- **полный набор данных для EDA и обучения моделей** [Georgia.zip](#)
- **комплект тестовых данных для инференса** [test_ecg.zip](#)



RUNNING...

Stop



Приложение для анализа сигналов ЭКГ и определения состояния здоровья человека

Загрузите ZIP-файл для обучения



Drag and drop file here

Limit 200MB per file • ZIP

Browse files



Georgia.zip 69.1MB



Разведочный анализ данных

Пожалуйста, загрузите файл, чтобы произвести анализ.

Обучение модели

Пожалуйста, загрузите файл, чтобы обучить модель.

Прогноз по анализам ЭКГ

Выберите модель для предсказания

17_42_30_12_2024_63fb0fa0



Информация о выбранной модели

20:19 Пн 30 дек.

Study


АА

Небезопасно — 195.133.13.244


66 %

Приложение для анализа сигналов ЭКГ и определения состояния здоровья человека

Загрузите ZIP-файл для обучения

 Drag and drop file here
Limit 200MB per file • ZIP

Browse files

 Georgia.zip 69.1MB

×

Файл успешно загружен!

Разведочный анализ данных

Загруженный файл: Georgia.zip

Провести EDA

Обучение модели

Выберите модель:

Модель

SVC

С

1.00

0.00

100.00

- Нажмите кнопку "Провести EDA".
- Подождите, пока файл загрузится и появится надпись "EDA успешно выполнен!".
- На экране появятся раздел "Описательная статистика". Используйте кнопки и фильтры для настройки отображения данных.

20:20 Пн 30 дек. AA Небезопасно — 195.133.13.244

Разведочный анализ данных

Загруженный файл: Georgia.zip

Провести EDA

Обучение модели

Выберите модель:

Модель

SVC

C

1.00

0.00 100.00

Kernel

linear

Gamma

scale

Class Weight

None

Обучить модель

Прогноз по анализам ЭКГ

Выберите модель для предсказания

17_42_30_12_2024_63fb0fa0



Разведочный анализ данных

Загруженный файл: Georgia.zip

Провести EDA

EDA успешно выполнен!

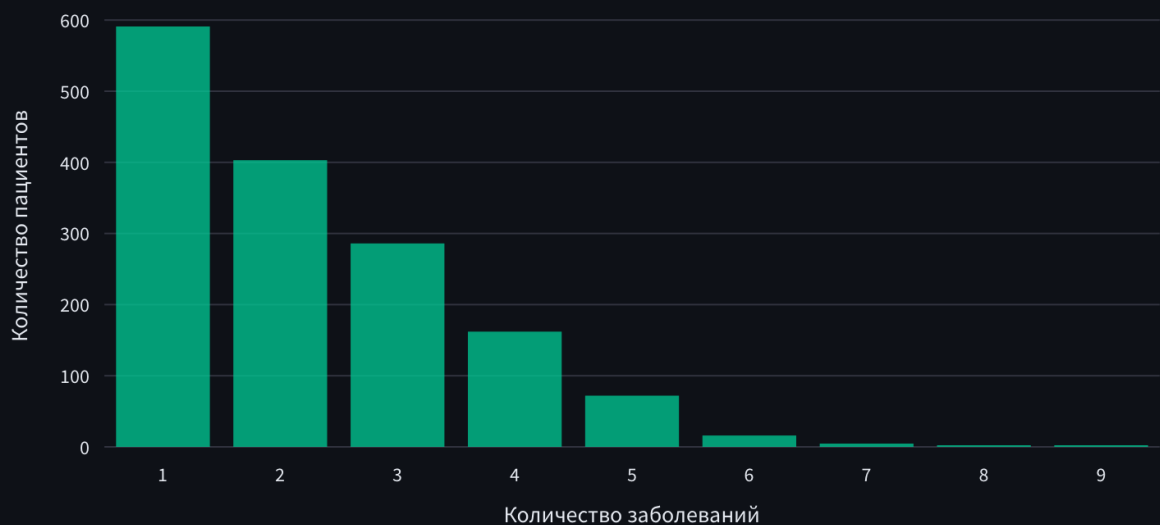
Описательная статистика:



	one	two	three	aVR	aVL	aVF	V1	V2	V3	V4
count	3,421	3,421	3,421	3,421	3,421	3,421	3,421	3,421	3,421	3,421
mean	-3.6726	0.6878	4.3978	1.4583	-4.0278	2.5498	-4.1476	0.9503	6.1137	7.1386
std	177.392	205.863	192.2008	166.357	153.622	178.2661	228.9045	309.6223	270.1557	300.9837
min	-1,356	-1,356	-1,268	-1,551	-973	-758	-2,464	-3,294	-2,762	-7,915
25%	-68	-78	-48	-19	-53	-58	-29	-53	-58	-78
50%	-29	-24	0	29	-14	-9	14	9	0	-9
75%	19	39	48	63	24	39	58	73	78	53
max	1,434	2,205	2,059	1,356	1,195	2,132	2,415	1,961	1,971	2,664



Гистограмма распределения количества заболеваний у одного пациента



5. Обучение модели В разделе "Обучение модели" :

- Выберите модель из предложенного списка и необходимые гиперпараметры
- Нажмите кнопку "Обучить модель" .
- Подождите, пока приложение отправит данные на сервер и получит результат.

20:21 Пн 30 дек. 66 %

Study < > AA Небезопасно — 195.133.13.244

⌵

Обучение модели

Выберите модель:

Модель

Logistic Regression

Solver

liblinear

Penalty

l1

C

0.001

Class Weight

None

Обучить модель

Прогноз по анализам ЭКГ

Выберите модель для предсказания

17_42_30_12_2024_63fb0fa0

Информация о выбранной модели

Метрики:

```
{  "accuracy" : 0.1873198847262248  "f1" : 0.0746547980104482 }
```

Параметры:

```
{
```

6. Анализ ЭКГ на основе выбранной модели Результаты отображаются в разделе "Прогноз по анализам ЭКГ" в удобном формате, для этого необходимо:

- Выберите модель для предсказания из предложенного списка, **обратите внимание на дату и время в названии файлов из предложенного списка.**
- Загрузите ZIP-файл для прогноза диагноза.

- Нажмите кнопку "Получить прогноз"
- Получите уведомление с ключевыми выводами.

20:21 Пн 30 дек. AA Небезопасно — 195.133.13.244

Study

Обучить модель

Прогноз по анализам ЭКГ

Выберите модель для предсказания

17_42_30_12_2024_63fb0fa0

Информация о выбранной модели

Метрики:

```
{
  "accuracy" : 0.1873198847262248
  "f1" : 0.0746547980104482
}
```

Параметры:

```
{
  "model" : "Logistic Regression"
  "solver" : "liblinear"
  "penalty" : "l1"
  "C" : "0.01"
  "class_weight" : "None"
}
```

Получение прогноза о состоянии здоровья

Загрузите ZIP-файл для прогноза диагноза

Drag and drop file here
Limit 200MB per file • ZIP

Browse files