

Проект MedAssistAI

Документация

Описание и структура проекта

Проект направлен на предсказание наличия сердечно-сосудистых заболеваний у человека (людей) по его (их) физическим параметрам. Проект состоит из двух частей:

- Клиентская часть (streamlit)
- Серверная часть (fastapi)

Серверная часть

Серверная часть состоит из 6 API эндпоинтов:

- /participants - выводит информацию об участниках проекта
- /model/fit - обучение модели на тренировочном датасете с возможностью загрузки желаемых гиперпараметров и максимальном времени исполнения запроса
- /model/set_model - опция установить обученную выше модель в пространство использования для предсказания. При запуске сервиса автоматически устанавливается заранее предобученная модель.

Серверная часть

- `/model/predict` - получение предсказаний на одном или множестве объектов данных моделью, которая установлена в пространство предсказаний. Возвращает json файл с информацией о probability предсказания.
- `/model/models` - опция, позволяющая вернуть список всех существующих обученных моделей с краткой информацией.
- `/model/update_model/{model_id}` - возможность переобучить существующую модель на новых данных

Модель

В нашем сервисе используется композиция из двух моделей - логистической регрессии и XGBClassifier. За подготовку данных и обучение логистической регрессии отвечает класс HeartBasedPredictor. За XGBClassifier - CardioTrainBasedPredictor. В сервисе же используется PredictorComposer, сериализованный в .pickle. В нем осуществляется композиция моделей и делегирование вызовов в предикторы. Обучалась дефолтная модель на данных, подготовленных на предыдущих чекпоинтах.

Клиентская часть

Интерфейс клиентской части состоит из боковой части слева, где можно либо загрузить табличку с данными о нескольких пациентах, либо выставить параметры для одного пациента вручную. Далее эта информация пойдет на вход моделям для получения предсказания о наличии или отсутствии заболевания.

Данные о пациенте
(пациентах) для постановки
диагноза

Загрузить CSV файл

Drag and drop file here

Limit 200MB per file • CSV

Browse files

Или выставите параметры вручную

Возраст

55

- +

Пол

Мужской

▼

Рост

170

- +

Вес

75

- +

Верхнее давление

120

- +

Клиентская часть

Так же есть основной блок, в котором отображается предварительная информация о параметрах пациента, и есть несколько вкладок с уникальным функционалом.

Данные пациентов

	Age	Sex	CheastPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngi
0	55	1	<NA>	130	240	<NA>	<NA>	150	<NA>

Fit ▾

Set model ▾

Predict ▾

Predict proba ▾

Models ▾

Update model ▾

Participants ▾

Клиентская часть

Вкладка fit содержит в себе функционал для обучения новых моделей на тренировочном датасете в формате CSV. Также необходимо предоставить JSON файл с описанием гиперпараметров. С помощью значения таймаут устанавливается максимальное время обучение модели. В поле ID нужно ввести желаемое название модели в произвольной форме.

Fit

Обучить модель с заданными гиперпараметрами на ваших данных

Введите ID новой модели

best_model_in_the_world_1

Загрузить CSV файл с обучающими данными

Drag and drop file here

Limit 200MB per file • CSV

Browse files

Загрузить JSON файл с гиперпараметрами

Drag and drop file here

Limit 200MB per file • JSON

Browse files

Введите время для таймаута (сек)

10

- +

Отправить запрос

Клиентская часть

Во вкладке Set model есть возможность установить модель для использования ее в предсказании по ее ID, который устанавливался на шаге выше. Для установки заранее предобученной модели используется имя ID default_model.

Set model

Установите модель в качестве активной для дальнейших предсказаний

Введите ID обученной модели

Press Enter to apply

Отправить запрос

Клиентская часть

С помощью вкладок Predict и Predict proba можно получить конкретное предсказание или вероятность относительно заболевания. Предсказания выполняются моделью, которая была установлена во вкладке Set model.

Predict

Предсказание на основе данных пациентов

Требуется заполнить sidebar: загрузить CSV файл или выставить параметры вручную

Отправить запрос

Predict proba

Вероятность сердечного заболевания на основе данных пациентов

Требуется заполнить sidebar: загрузить CSV файл или выставить параметры вручную

Отправить запрос

Клиентская часть

Во вкладке Models есть возможность получить информацию о существующих моделях. Здесь есть как и обученные модели с помощью fit, так и предобученная заранее модель. По каждой модели выводится информация о ее ID, времени создания, флаг нахождения в пространстве предсказаний и словарь гиперпараметров.

Models

Получение списка всех доступных моделей и информации о них

Отправить запрос

Запрос отправлен!

Доступные модели

	id	created_at	is_active	parameters
0	default_model	2024-12-31T03:58:07.897558	true	{}
1	model1	2024-12-31T04:46:17.897706	false	{"learning_rate":0.01,"n_estimators":1

Клиентская часть

Во вкладке Update model можно переобучить существующую модель на новом тренировочном датасете. Для этого необходимо указать модель в поле ID и загрузить CSV таблицу с новыми данными.


Update model

Обновление существующей модели новыми данными

Введите ID модели для обновления

best_model_in_the_world_1

Загрузить CSV файл с новыми тренировочными данными

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

Browse files

Отправить запрос