



УНИВЕРСИТЕТ
ЛОБАЧЕВСКОГО



Разработка детектора ковидных
аномалий в ритме сердца

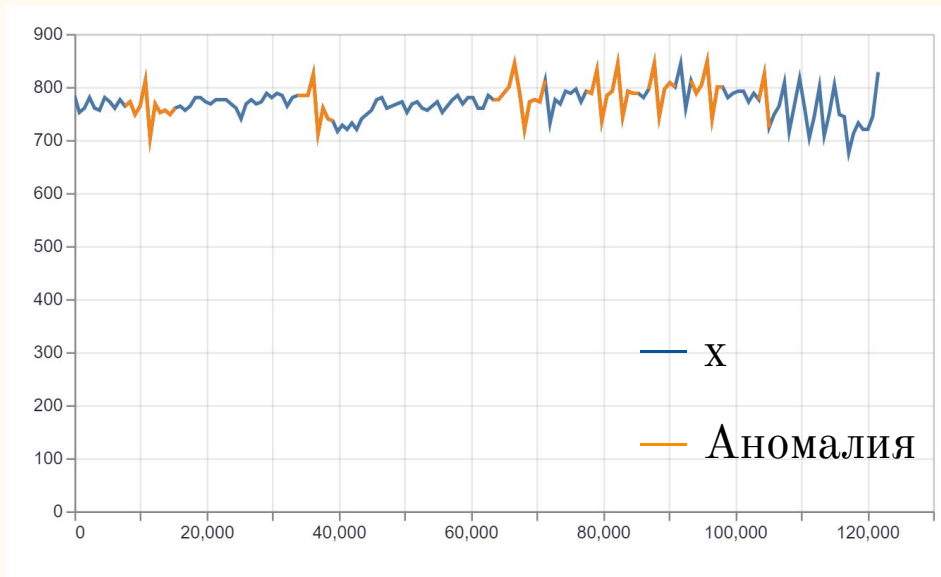
Проблема

Быстро и точно детектировать проблемы с сердцем,
возникающие вследствие COVID-19



Решение основной проблемы

Построение качественного детектора аномалий в ритме сердца для своевременного оповещения человека об изменении сердечного ритма, связанного с коронавирусом

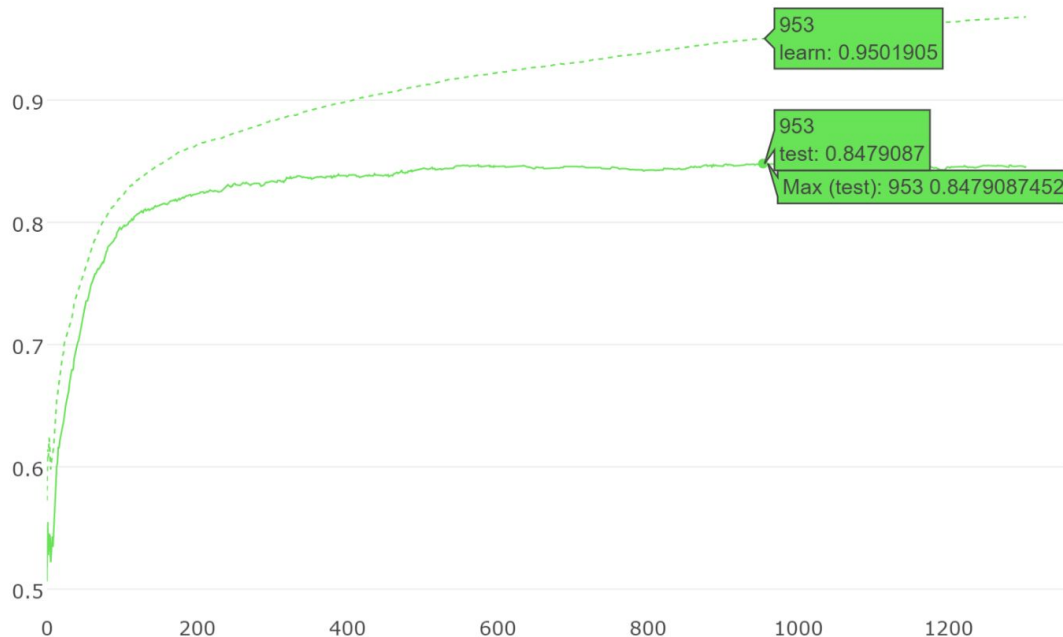


Решение основной проблемы

☒ Learn ☒ Eval
☒ catboost_info ~1s 821ms 32s 972ms
 ---- learn ---- test
 curr ---- 0.9501905... 0.8479087... 953
 best 0.8479087... 953

F1 Logloss

Кривая обучения



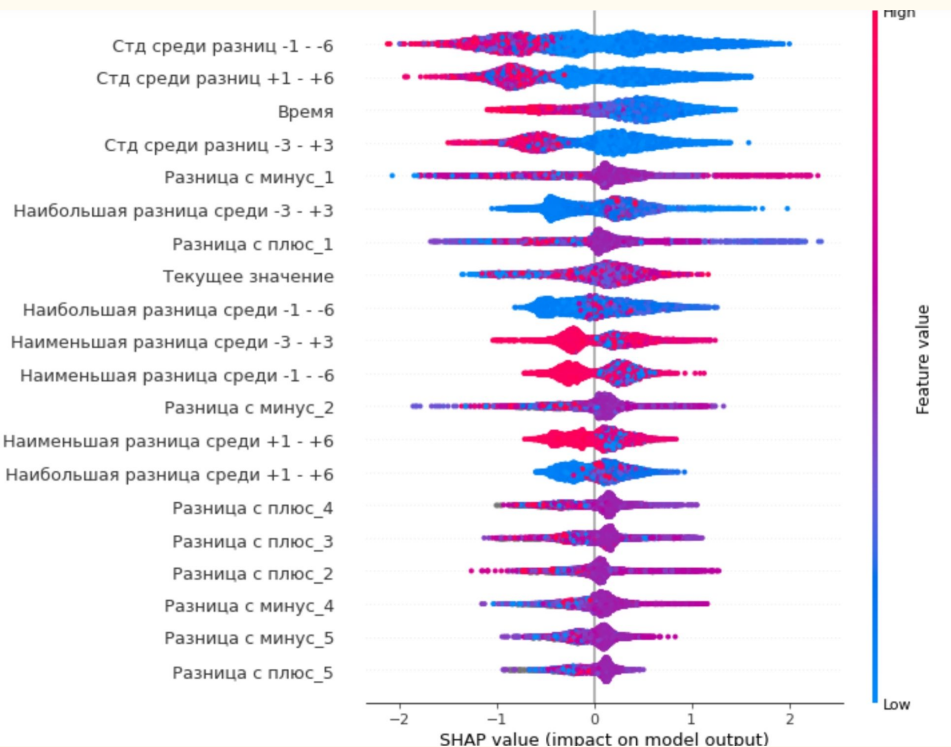
Модель Catboost

Признаки —
 разности значения точки
 с соседними точками
 в некотором окне

☐ Click Mode ☐ Logarithm
☐ Smooth 0

Решение

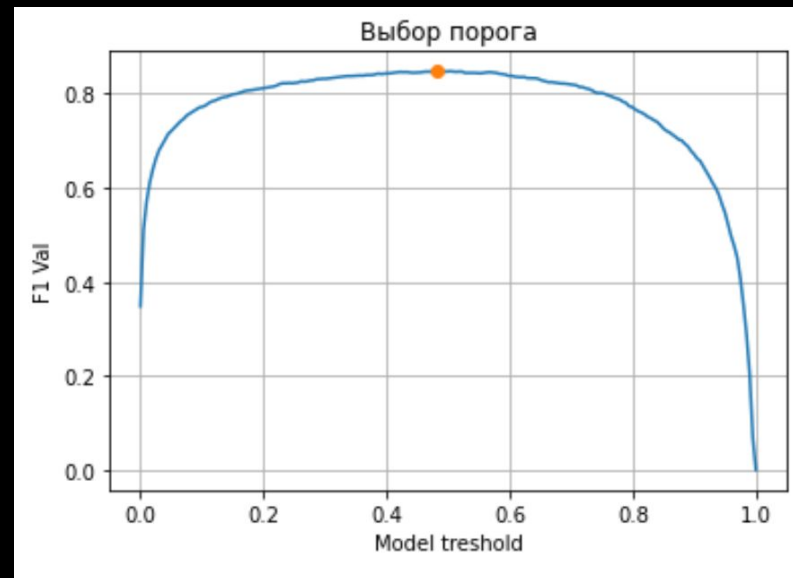
Признаки по убыванию значимости



Модель Catboost



Признаки —
разности значения точки
с соседними точками
в некотором окне



Новаторские идеи

- Решение на стыке медицинской и технической наук
- Максимально увеличить скорость принятия решения моделью
- Использование аугментации данных
- Детектирование и устранение выбросов

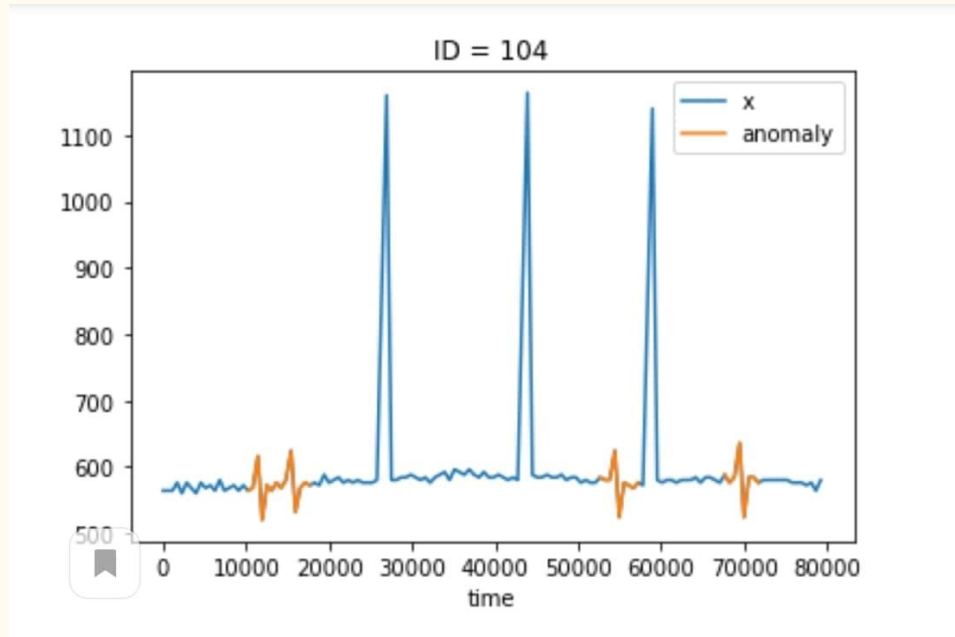
Второстепенные проблемы

- Необходимость заполнить шумы
- Расширить выборку открытыми данными (kaggle, mendeley data)

BIG DATA

Аналитика

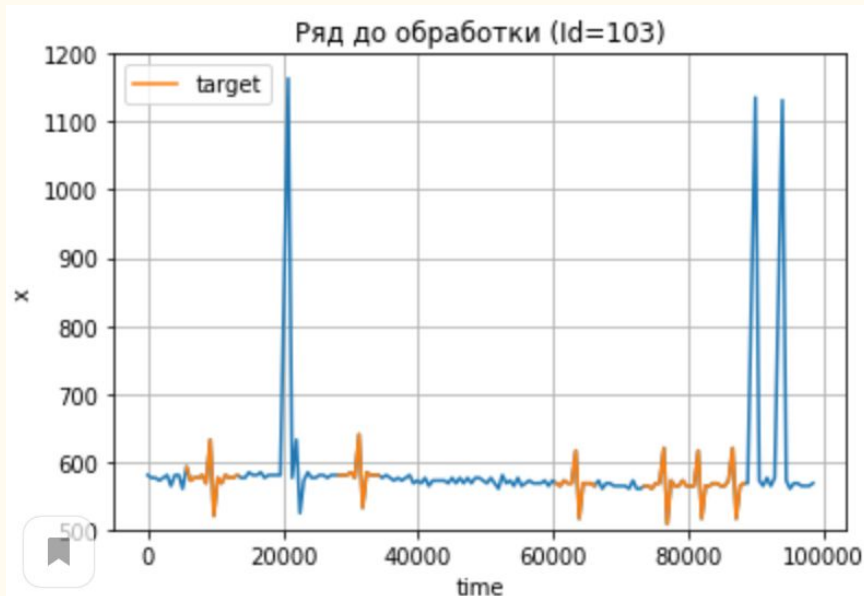
- Аномалии цикличны и имеют схожий характер для одного человека
- Есть выбросы, которые можно связать с пропуском удара пульса из-за недостаточного прилегания датчика или чего-то подобного



Работа с данными

Очистка от выбросов (hrv-analysis, детектирование с помощью созданных фичей)

Создание и отбор разных признаков (скользящее среднее, агрегационные функции, разности)



Работа с данными

Сгенерированный текст

с разной температурой:

IJIIJIIINGLGHIIIIIN

IJIIJIIINGLGHIIIIIN

JIIJIIJGNNHGEEDDDDE

IJIIJIIJGNNHGLGIIIIIN

JIDJIIJNJILLIIIIJIL

Оригинал-----

DDCCCCDDGGIIDVBCCD

Цикличность RR интервала

- Кодирование интервалов интервал-буква между квантилями
- Нормирование данных на медиану

Генератор

Рекуррентная нейронная сеть LSTM

Демо решения

Ссылка на работающее демо:

<https://clck.ru/VbovE>



Какие данные скорить?

Тестовый датасет

Скоринг завершён успешно!

[Скачать результаты в формате csv](#)

Доля класса 1 = 10.8 %

Результат работы модели для различных ID

Выберите id (сортировка в порядке уменьшения числа выявленных аномалий)

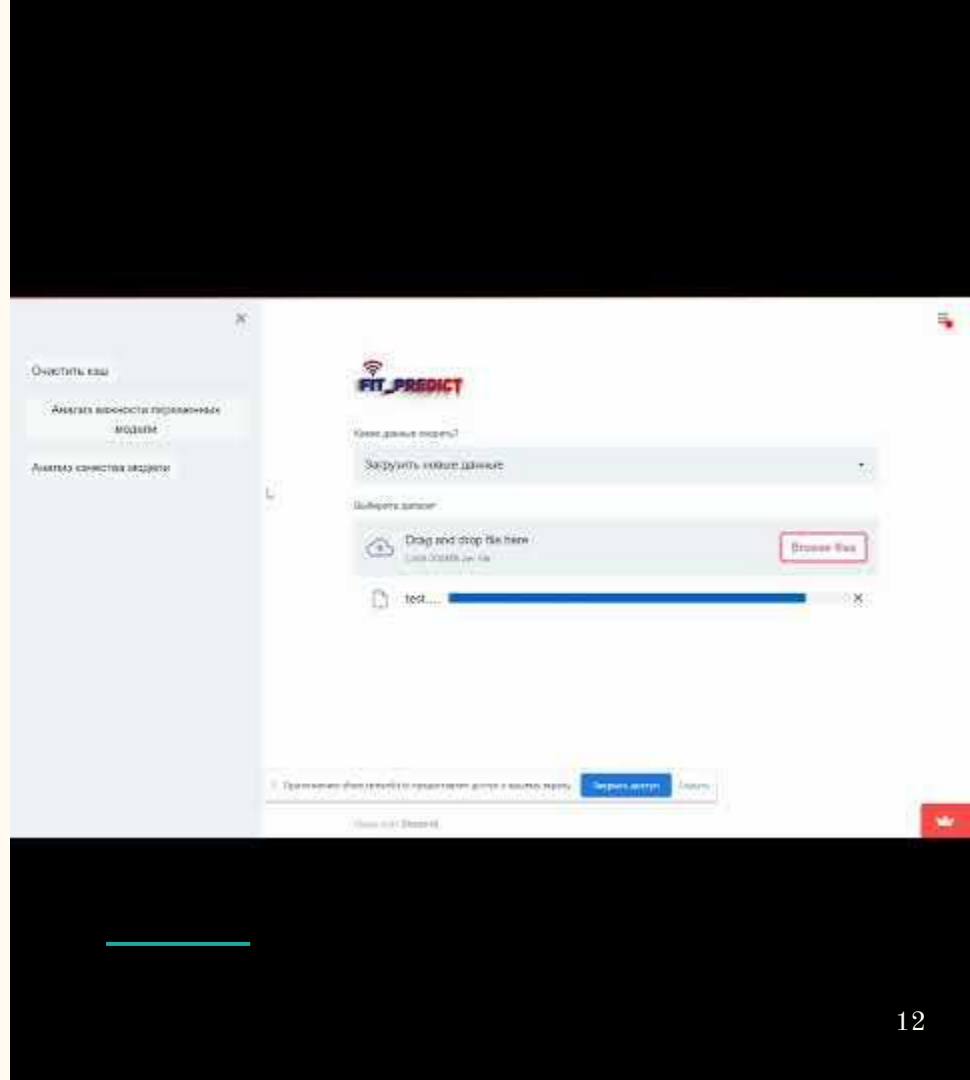
106



Демо решения

Ссылка на работающее демо:

<https://clck.ru/VbovE>



Стек задействованных технологий

Итоговое решение

Модели Catboost и Lightgbm показали одинаковый скор на приватном лидерборде

Модель логистической регрессии показывала значимо меньший скор

Программные Инструменты

- Jupyter notebook, Python



Машинное обучение

- Shap (отбор значимых признаков)



- CatBoost, LightGBM

(построение моделей)



- Optuna

(оптимизация гиперпараметров моделей)



Фронтенд:

- Streamlit



Наша команда



С благодарностью за Ваше внимание!

Пермь



Data Science

Олег Черемисин

@mktoid

cheremisin@gmail.com

Москва



Data Science

**Альбина
Ахметгареева**

@Colindonolwe

albina.akhmetgareeva@gmail.com

Москва



Data Science

Дима Васькин

@vaskind

vaskind@vandex.ru

