

Хакатон Cardio Spike

Команда Звездочка
Сервис "Сердечный друг"

Постановка задачи

- Диагностика аномалий сердечных ритмов, специфичных для больных COVID-2019 пациентов - заблаговременное выявления сердечных осложнений
- Мало экспертов, обладающих навыками выявления таких аномалий на ритмограммах
- У большого количества людей осложнения возникают бессимптомно, но в дальнейшем представляют существенную опасность для их жизни и здоровья
- Входные данные - 300 R-R ритмограмм, около 6800 измерений
- Метрика качества - бинарная F1 мера



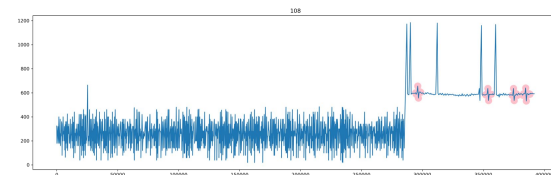
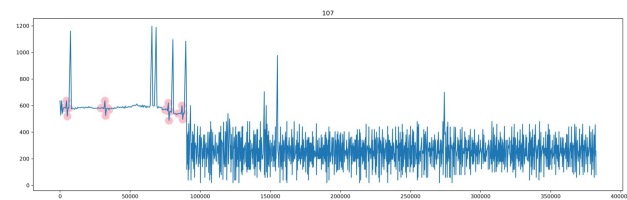
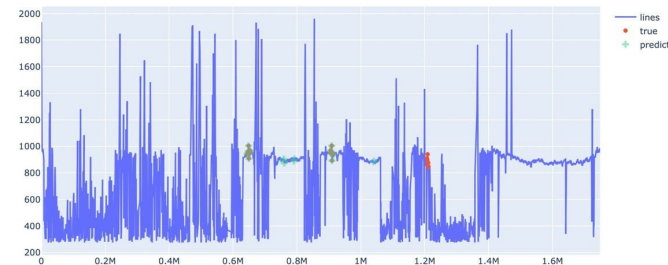
Вызовы задачи

- Сравнительно небольшой датасет
- Дисбаланс классов - аномалий в 8 раз меньше, чем без аномальных участков
- “Разрезанные” измерения
- Шум и ошибки в измерениях
- Ручная разметка без дополнительной валидации



Подготовка данных

- Удаление экстремальных пиков и провалов (> 300 и < 1400)
- Преобразование датасета в "окна", целевая точка в центре "окна"
- Масштабирование значений в окнах
- Расчет статистик и медицинских параметров для окна и датасета:
 - RMSSD - квадратный корень среднего из квадратов разниц последовательных интервалов
 - SDNN - стандартное отклонение интервалов
 - И другие
- Не помогло
 - Построение wavelets и спектрограмм - скорее всего из-за природы сетки
 - Предобработка размеченных аномалий - привело к ложным срабатываниям



Описание решения

- Композиция нескольких моделей
 - LightGBM
 - Нейронные сети: CNN + RNN + Attention
 - Ridge для композиции решений
- Подбор гипер параметров для бустинга и нейронных сетей
 - Optuna + Hydra
 - На 4 серверах, распределенные в рамках одного эксперимента - подбор архитектуры и параметров нейронной сети



Проверка результата (валидация)

Для получения стабильного балла на ЛБ

- Двойная валидация
- Ансамбль моделей

The table illustrates a validation process across 30 rows (1-30) and 26 columns (A-Z). The columns are organized as follows:

- Column A:** Labeled 'ДАТАСЕТ' (Dataset).
- Columns B-C:** Labeled 'FOLD 1' and 'FOLD 2' respectively.
- Columns D-E:** Labeled 'TRAIN'.
- Columns F-G:** Labeled 'VALID'.
- Columns H-I:** Labeled 'ТУТ ВАЛИДИРУЕМ' (Here we validate).
- Columns J-K:** Labeled 'VALID'.
- Columns L-M:** Labeled 'TRAIN'.
- Columns N-O:** Labeled 'VALID'.
- Columns P-Q:** Labeled 'ТУТ ВАЛИДИРУЕМ'.
- Columns R-S:** Labeled 'VALID'.
- Columns T-U:** Labeled 'TRAIN'.
- Columns V-W:** Labeled 'VALID'.
- Columns X-Y:** Labeled 'ТУТ ВАЛИДИРУЕМ'.
- Column Z:** Labeled 'VALID'.

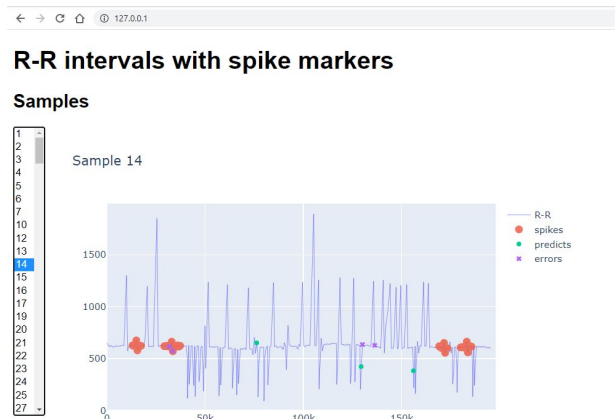
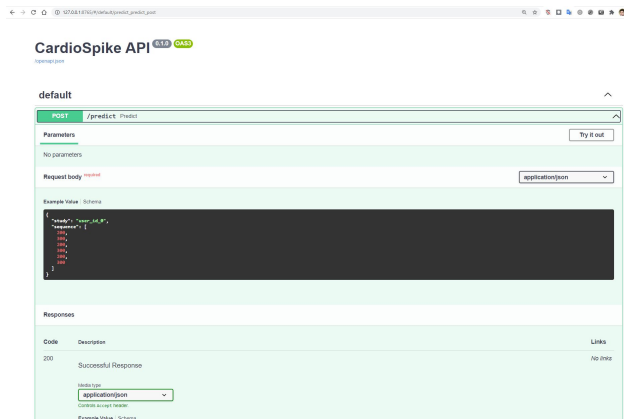
Each cell contains text indicating the status of the validation, such as 'VALID', 'TRAIN', 'ТУТ ВАЛИДИРУЕМ', or 'результат для стабильности' (stability result).

Результат на ЛБ:

- Финальный: 0.86 - 5-6 место

Демонстрация API и Интерфейса

- Интересный факт
 - Собственные данные собранные членами команды
 - Welltory - оптические измерения с использованием DN алгоритмов
 - Некоторые переболели ковид



Внешние данные и статьи

- [Репозиторий Welltory \(Связь HRV и Covid-19\)](#)
- [Предобработка несбалансированных RR интервалов](#)
- [Классификация RR интервалов](#)
- [Данные по HRV](#)
- [Применение визуального анализа RR для диагностики](#)
- [Wavelets для анализа RR интервалов](#)



Стек технологий - Open Source

Python, Optuna, Hydra, Pytorch, Pytorch Lightning, Sklearn, Pandas, Numpy, Plotly, Flask, FastAPI, Poetry





СБЕР МАРКЕТ

Спасибо от команды "Звездочка"!

Сергей
Lead ML



Даниил
ML



Глеб
Lead ML



Михаил
ML

