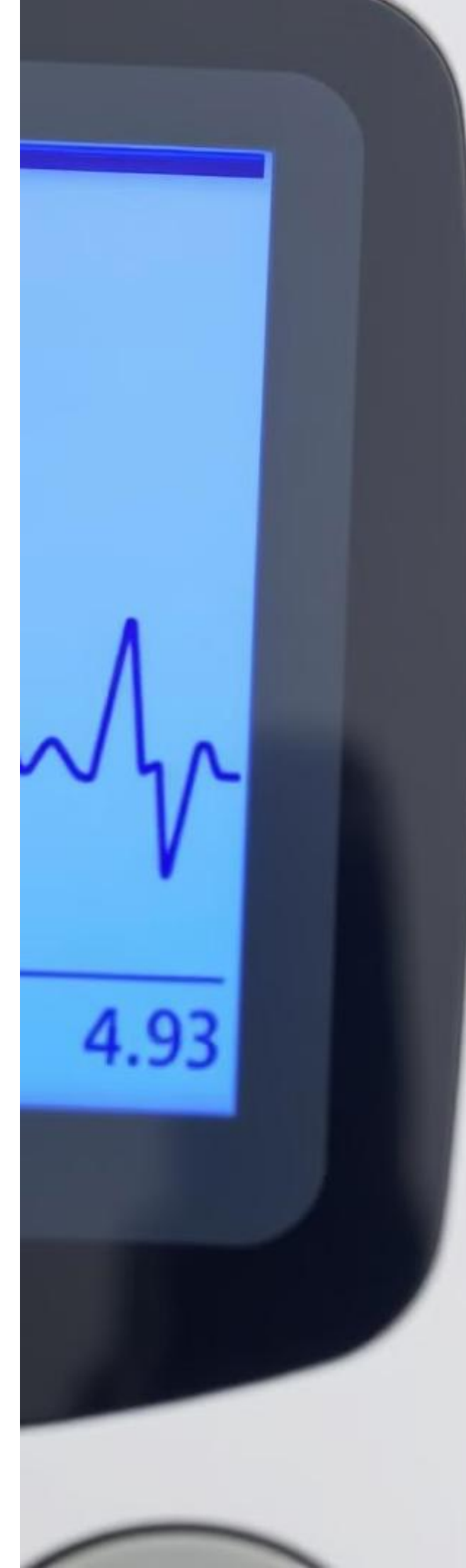


# Прогнозирование результатов ЭКГ: Выбор лучшей модели

В этой презентации мы рассмотрим процесс выбора и сравнения моделей машинного обучения, примененных для предсказания результатов ЭКГ. Будут представлены этапы предобработки данных, подбора оптимальных параметров моделей и анализа их метрик.



# Используемые модели

## CatBoostClassifier

Градиентный бустинг,  
известный своей  
устойчивостью к выбросам и  
высокой точностью

## LGBMClassifier

Еще один алгоритм  
градиентного бустинга,  
отличающийся быстротой  
обучения и хорошей  
обобщающей способностью

## RandomForestClassifier

Ансамбль решающих  
деревьев, эффективный при  
работе с большим  
количеством данных и  
высокой размерностью

## DecisionTreeClassifier

Простая модель, которая  
строит дерево решений для  
классификации, позволяющая  
легко интерпретировать  
результаты

## XGBoost

Оптимизированная  
реализация градиентного  
бустинга и разновидность  
метода коллективного  
обучения

# Предобработка данных

- 1 Из всех доступных заболеваний были выбраны 50 наиболее частых.
- 2 Пол пациента кодировался двоичным значением:  
Male = 0, Female = 1.
- 3 Проведены преобразования категориальных признаков, например, преобразование в числовые значения или использование one-hot encoding.
- 4 В результате анализа важности признаков были выбраны 20 наиболее значимых для прогнозирования.

# Сравнение метрик

Модель	Precision (Weighted)	Recall (Weighted)	F1-score (Weighted)
CatBoostClassifier	0.945	0.958	0.945
LGBMClassifier	0.950	0.959	0.944
RandomForestClassifier	0.884	0.907	0.871
DecisionTreeClassifier	0.856	0.870	0.862
XGBoost	0.514	0.310	0.320



# Вывод: LGBMClassifier — лучший выбор

Модель LGBMClassifier + MultiOutputClassifier продемонстрировала лучшие результаты по всем метрикам, особенно по F1-score, что указывает на ее сбалансированность и высокую точность.