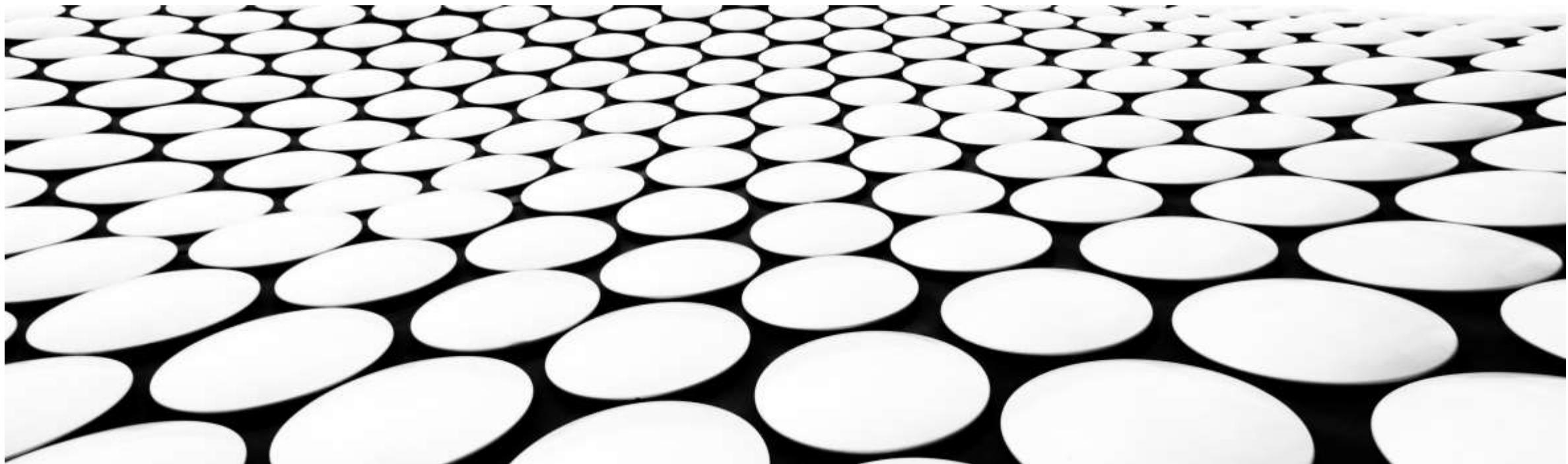

РАЗРАБОТКА СИСТЕМЫ АНАЛИЗА МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ ДЛЯ ЭПИДЕМИОЛОГИЧЕСКОГО МОНИТОРИНГА COVID-19

М24-525 КОЧЕТКОВ МИХАИЛ



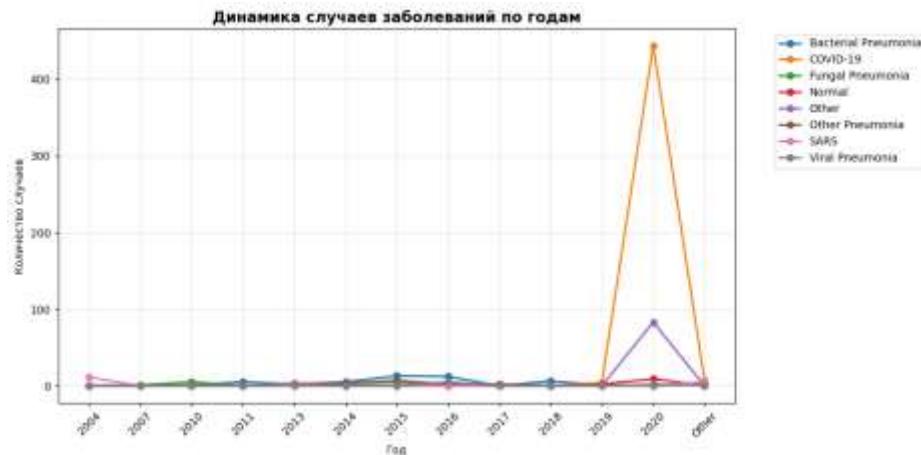
АРХИТЕКТУРА ПРОЕКТА

- 1. Источник данных
 - - Датасет 'COVID-19 Chest X-Ray Dataset' из открытого репозитория (<https://github.com/ieee8023/covid-chestxray-dataset>).
 - - Основной файл данных: 'metadata.csv', содержащий информацию о пациентах, диагнозах, проекциях снимков и временных метках.
- 2. Инструменты анализа:
 - - PySpark: используется для обработки больших объемов данных, выполнения SQL-запросов и агрегации метрик.
 - - Pandas: применяется для дополнительной обработки данных и подготовки их к визуализации.
 - - Matplotlib и Seaborn: библиотеки для создания визуализаций (графиков, тепловых карт, диаграмм).

ОСНОВНЫЕ СТАТИСТИКИ

В датасете доминируют случаи COVID-19 (61.47% всех диагнозов), средний возраст пациентов — 56.3 года. Наблюдается гендерный дисбаланс: 64% пациентов — мужчины, особенно заметный в группах с COVID-19 и пневмониями. Временные тренды показывают резкий рост числа исследований в 2020 году, где COVID-19 составил 82.53% всех случаев, что указывает на влияние пандемии.

АНАЛИЗ ДИНАМИКИ ЗАБОЛЕВАЕМОСТИ ПО ГОДАМ



Анализ заболеваемости по годам выявил характерную структуру: до 2020 года доминировали бактериальные пневмонии.

В 2020 году произошел резкий рост случаев — более чем в 20 раз — и кардинальное изменение структуры: COVID-19 стал абсолютным лидером, составив 82.5% всех случаев.