**Модель для определения эмоциональной окраски речи.**

**Цель:** Распределение аудио-сообщений по эмоциональной окраске для дальнейшего использования в качестве части дискриминатора в модели генерации речи.

**Задачи:**

1. анализ существующих решений;
2. сбор данных и их обновление;
3. обучение и оценка моделей на исходном датасете;
4. выбор модели с наибольшим значением точности;
5. оценка качества и скорости работы модели на новых аудио-сообщениях путем A / B тестирования;
6. выбор и развертывание наилучшей модели;
7. оптимизация выбранной модели путем квантизации и дистилляции;
8. развертывание оптимизированной модели, количественная оценка эффекта оптимизации.

**Датасет:** [Sber Golos. Датасет русской речи.](https://github.com/sberdevices/golos)

**Проект:** [sikoraaxd/speech-emotion-recognition](https://github.com/sikoraaxd/speech-emotion-recognition)

**Целесообразность использования датасета для решения поставленной задачи:** Golos - это датасет, собранный и размеченный компанией Sber, состоящий из 1240 часов или более одного миллиона коротких записей речи на русском языке. Датасет разделён на 2 части - записи с небольшого расстояния до устройства записи (Crowd) и большого растояния до устройства записи (Farfield) Каждая запись была проверена несколькими людьми, для маскимального избежания ошибок. Данную модель можно использовать для обучения моделей распознавания и синтеза речи.

**Обработка датасета:** Для достижения поставленной цели датасет необходимо дополнить вспомогательной информацией, а именно - информацией об эмоциональной окраске каждой записи. В рамках лабораторной работы для экономии времени разметка будет проводиться на выборке из данных, необходимой для достижения высокой точности распределения.

**Практическая значимость:** Модель будет использоваться как часть дискриминатора при генерации речи обучающей моделью. Дискриминатор будет штрафовать варианты генератора, не подходящие по эмоцианальной окраске. Таким образом, данная модель поможет добавить экспрессивности к генерируемой речи.

**Краткая информация о датасете:**

На рисунке 1 показано распределение количества записей (ось Y) в зависимости от их длительности в секундах (ось X).

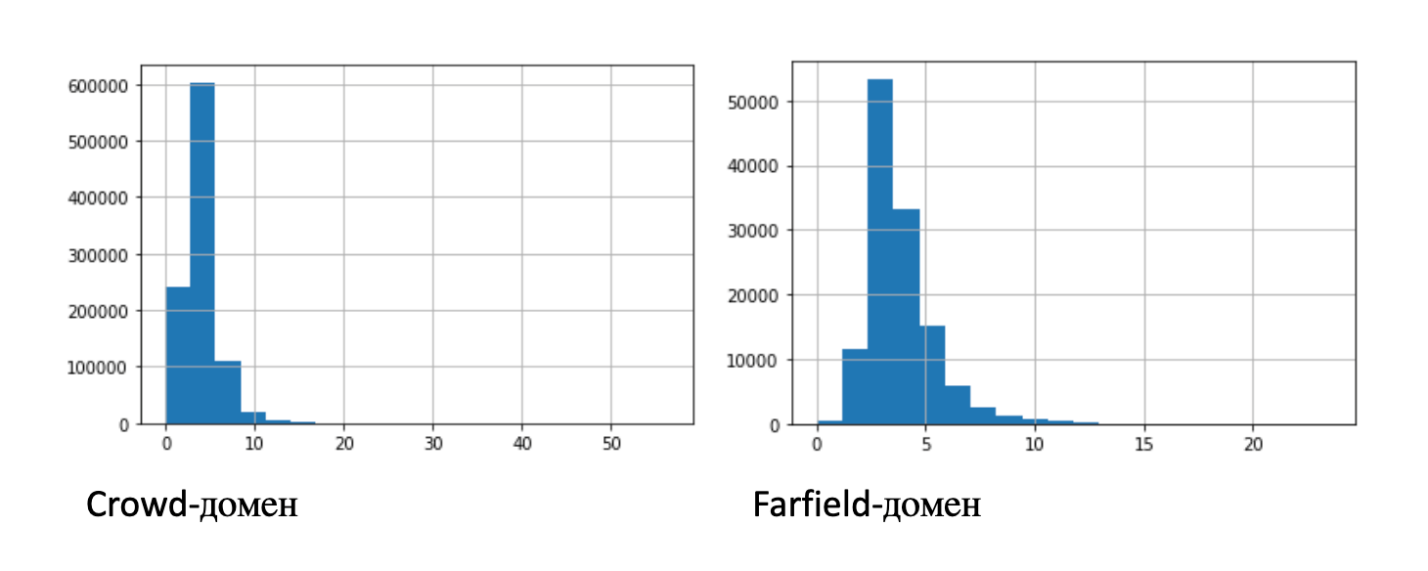
[](https://github.com/sikoraaxd/speech-emotion-recognition/blob/master/assets/charts/duration-count-hist.png)

Рисунок 1 - Распределение количества записей от их длительности

***Статистические данные:*** количество данных, средняя продолжительность записи, стандартное отклонение и перцентили.

[](https://github.com/sikoraaxd/speech-emotion-recognition/blob/master/assets/charts/dataset-stats.png)

Рисунок 2 - Статистические данные