**Задание 1.**

**Таблица 1 – Сравнение моделей для распознования текста**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Модель | Архитектура | Точность (WER) | Скорость распознования | Поддержка языков | Особенности |
| Wav2Vec2 | Transformers | ~4.8% | Высокая | Многоязычная поддержка | Высокая точность, требует мощного оборудования |
| DeepSpeech | RNN | ~7.2% | Средняя | Ограниченная поддержка | Поддержка низкопроизводительных  устройств |
| Google Speech-to-Text | RNN + CNN | ~4.9% | Высокая | Многоязычная  поддержка | Облачное решение, высокая точность |

**Обоснование выбора модели.**

Исходя из анализа, Wav2Vec2 выглядит как лучший выбор для большинства приложений, требующих высокую точность и скорость распознавания. Модель Wav2Vec2 от Facebook AI Research (FAIR) демонстрирует низкий уровень WER и быстрое распознавание благодаря архитектуре Transformer. Она также поддерживает множество языков, что делает её гибким инструментом для различных задач.

Однако, если критическим фактором является использование облачных технологий и широкая поддержка языков с готовыми API, то Google Speech-to-Text также будет отличным выбором.

**Задание 2.**

### **Анализ и обоснование выбранных методов и моделей**

Выбранные методы:

1. **Wav2Vec2 для распознавания речи**
2. **MFCC для извлечения признаков аудиофрагментов**
3. **KMeans для кластеризации фрагментов по спикерам**

Анализ методов

1. **Wav2Vec2 для распознавания речи**
   * **Описание**: Wav2Vec2 - это модель глубокого обучения на основе трансформеров, разработанная для задач автоматического распознавания речи (ASR). Модель обучается на сырых аудиоданных и имеет возможность захватывать сложные зависимости в аудиосигнале.
   * **Преимущества**:
     + Высокая точность распознавания речи благодаря использованию трансформеров.
     + Возможность работать с многими языками, включая русский.
     + Не требует ручного извлечения признаков.
   * **Недостатки**:
     + Требует значительных вычислительных ресурсов для работы и обучения.
     + Модель может быть чувствительна к шумам и качеству записи.
2. **MFCC (Мел-частотные кепстральные коэффициенты) для извлечения признаков**
   * **Описание**: MFCC - это распространенный метод извлечения акустических признаков из аудиосигнала. Он используется для представления спектральной информации в более компактной форме, которая хорошо подходит для анализа звука.
   * **Преимущества**:
     + Широко используется и доказал свою эффективность в различных задачах обработки речи.
     + Хорошо подходит для моделирования человеческого восприятия звука.
     + Относительно прост в вычислении.
   * **Недостатки**:
     + Может не захватывать некоторые важные временные зависимости.
     + Чувствителен к шумам и искажениям аудиосигнала.
3. **KMeans для кластеризации фрагментов по спикерам**
   * **Описание**: KMeans - это алгоритм кластеризации, который группирует данные в k кластеров на основе их признаков. В данном контексте используется для разделения фрагментов аудио по разным спикерам.
   * **Преимущества**:
     + Простой и быстрый алгоритм для выполнения кластеризации.
     + Хорошо подходит для задач, где количество кластеров заранее известно или может быть оценено.
   * **Недостатки**:
     + Количество кластеров должно быть задано заранее.
     + Может плохо работать на сложных данных с нерегулярной структурой кластеров.

#### Обоснование выбора методов и моделей

1. **Wav2Vec2** была выбрана из-за её высокой точности и производительности в задачах распознавания речи. Модель эффективно работает с аудиоданными и обеспечивает высокое качество транскрипции даже для сложных и шумных аудиозаписей.
2. **MFCC** использован для извлечения признаков из аудиосигнала, так как это один из наиболее популярных и проверенных методов в области обработки речи. Он предоставляет информативные признаки, которые хорошо подходят для последующей кластеризации.
3. **KMeans** был выбран для кластеризации фрагментов по спикерам благодаря своей простоте и эффективности. Алгоритм позволяет быстро и эффективно группировать аудиофрагменты, что важно для задач с большими объемами данных. При этом его недостатки (такие как необходимость заранее задавать количество кластеров) могут быть компенсированы предварительным анализом данных или использованием эвристик.