

## Отчет

В этой серии заданий поставила себе задачи, чтобы узнать и попробовать возможности различных моделей и направлений в машинном обучении. В итоге были поставлены следующие задачи:

### Задача 1. Фильтр комментариев

Цель: классификация комментариев на позитивные и негативные, вывод статистики и сохранение всех комментариев в файл.

Инструменты: transformers, модель ruBERT-base-cased-sentiment-rureviews.

Вход: список текстовых комментариев.

Выход: статистика (количество и процент положительных/отрицательных комментариев); файл comments\_result.txt со всеми комментариями.

### Задача 2. Подсказки текста при ограничении символов

Цель: если текст превышает лимит символов тогда нужно сгенерировать несколько сокращённых вариантов текста пользователя.

Инструменты: transformers, модель ruT5-base-summarizer.

Вход: Введенный пользователем текст и ограничение по символам.

Выход: 3 варианта сокращенных текстов для выбора пользователем.

### Задача 3. Голосовые команды для умного дома

Цель: распознавание речи и выполнение команд управления (свет, музыка, температура, двери, увлажнитель воздуха и тд).

Инструменты: transformers, модель whisper-small.

Вход: аудиофайл (wav/mp3)

Выход: Распознанная команда выполняется (например, свет включен/выключен).

#### Задача 4. Распознавание объектов на изображении

Цель: определить изображение на картинке.

Инструменты: torchvision, модель ResNet-50 (ImageNet).

Вход: путь к изображению URL.

Выход: название предсказанного объекта.

#### Задача 5. Распознавание действий на видео

Цель: Определить действие происходящее в коротком видеоролике.

Инструменты: torchvision, модель MC3\_18 (Kinetics-400).

Вход: путь к видеофайлу.

Выход: список из трех наиболее вероятных действий с вероятностью совпадения.

#### Преимущества выбранных библиотек

- 1) Transformers (Hugging Face) - простая интеграция готовых NLP и ASR моделей, большая библиотека предобученных моделей, удобный pipeline API.
- 2) Torch/Torchvision - гибкость в построении и обучении моделей, поддержка GPU и CUDA, большой набор предобученных моделей для изображений и видео.
- 3) PyAv - удобное чтение и обработка кадров из видео.

#### Принцип работы моделей

- 1) RuBERT (Sentiment): трансформерная архитектура BERT, адаптированная под русский язык, высокая точность на тональности текста.
- 2) ruT5 (Summarizer): модель типа Encoder-Decoder, подходит для генерации сокращённых текстов, оптимально для задачи сокращения.

- 3) Whisper (ASR): универсальная модель распознавания речи от OpenAI, работает стабильно на разных языках.
- 4) ResNet-50: сверточная нейросеть с остаточными блоками, стандарт для классификации изображений.
- 5) MC3\_18: 3D-CNN модель, которая учитывает как пространственные, так и временные признаки, что важно для видео.

## Причины выбора

- 1) Предобученность моделей – все выбранные решения основаны на уже обученных архитектурах (BERT, T5, Whisper, ResNet, MC3), что значительно сокращает время разработки и избавляет от необходимости собирать и размечать большие датасеты.
- 2) Поддержка русского языка – для задач обработки текста и речи выбраны модели, которые адаптированы или обучены на русскоязычных данных (rubert, ruT5, Whisper), что повышает качество распознавания и генерации.
- 3) Широкая поддержка сообществом – все модели доступны в Hugging Face или Torchvision, активно поддерживаются и обновляются. Это снижает риски использования устаревших решений.
- 4) Универсальность и проверенная эффективность – модели, такие как ResNet-50 и MC3, являются стандартом де-факто для классификации изображений и видео и показывают стабильные результаты в разных приложениях.
- 5) Интеграция с инфраструктурой PyTorch – все модели легко запускаются как на CPU, так и на GPU, что позволяет масштабировать решения под разные сценарии (от локального запуска до облачного сервиса).
- 6) Гибкость дообучения – при необходимости их можно дообучить (fine-tuning) под конкретный бизнес-кейс без полной тренировки с нуля (это было бы очень удобно для 2 задачи, так как там модель

достаточно глупа, потому что было обучена не совсем на тех данных, которые мне были нужны).

- 7) Высокая воспроизводимость и документация – каждая модель имеет описанную структуру датасета, метрики качества и примеры использования, что облегчает тестирование и интеграцию.

### Структура датасета

- 1) Sentiment - датасет из отзывов на русском языке (позитивные/негативные).
- 2) Summarization - тексты и их сокращённые версии из различных газет.
- 3) ASR - аудиозаписи с транскрипцией.
- 4) ImageNet (ResNet) - 1.2 млн изображений, 1000 классов.
- 5) Kinetics-400 - ~400 тыс. видеороликов, 400 классов действий.

### Метрики и оценки качества моделей

- 1) ML-метрики: Sentiment → Accuracy, F1-score, Summarization → ROUGE, BLEU, ASR → WER (Word Error Rate), Image classification → Top-1 Accuracy, Top-5 Accuracy, Video recognition → Top-k Accuracy.
- 2) Продуктовые метрики: время ответа модели, удовлетворенность пользователя (например, по фидбеку), количество ошибок в реальных сценариях.
- 3) Инфраструктурные метрики: использование GPU / CPU, время инференса, потребление памяти.

### Особенности реализации и эффективность

Использование предобученных моделей позволило быстро решить задачи без долгого обучения. Для больших моделей (Whisper, T5, ResNet-50) важна работа на GPU, иначе время инференса сильно возрастает. Для Summarization введена генерация нескольких

вариантов текста, что повышает удобство для пользователя. Для Video recognition было оптимизировано чтение кадров (выбор равномерных num\_frames), что ускоряет предсказания.

## Локальная LLM

Так же была развернута локальная модель, был выбор между LmStudio и ollama, прочитала плюсы и минусы у одной и второй, в итоге выбрала ollama, сначала скачала программный интерфейс, скачала туда модель gemma:latest 3.3GB, ввела промт и LLM ответила. Открыла документацию на гитхабе и повторила все тоже, но уже через командную строку, все работало так же, увидела в документации, что можно запустить скрипт написанный на питоне используя эту локальную модель, написала простенький скрипт и выполнила его. Изучила названия моделей, какие есть, мне понравилась модель Qwen, почитала про нее и установила себе Qwen2.5vl:7b 6GB, так как мало памяти на ПК не смогла скачать модели помощнее.