

ГНУ MUSIC

ML



Задача

Креативный фильтр

Команда



**Кирилл
Рыжичкин**

AI/DL Specialist

@polnostju



**Александр
Груздев**

AI/DL Specialist

@TheStrangerOne



**Владимир
Москвин**

AI/DL+ Analytics

@virtuozm



**Егор
Миронов**

Front/Back-end'er

@voultem

Кейс

Цель

Создание **инструмента**, который сможет отслеживать креативы брендов и на основании наполнения аудио/видеофайлов сможет определить **сегментацию** — в какую категорию определить медиа давление.

Требования к модели:

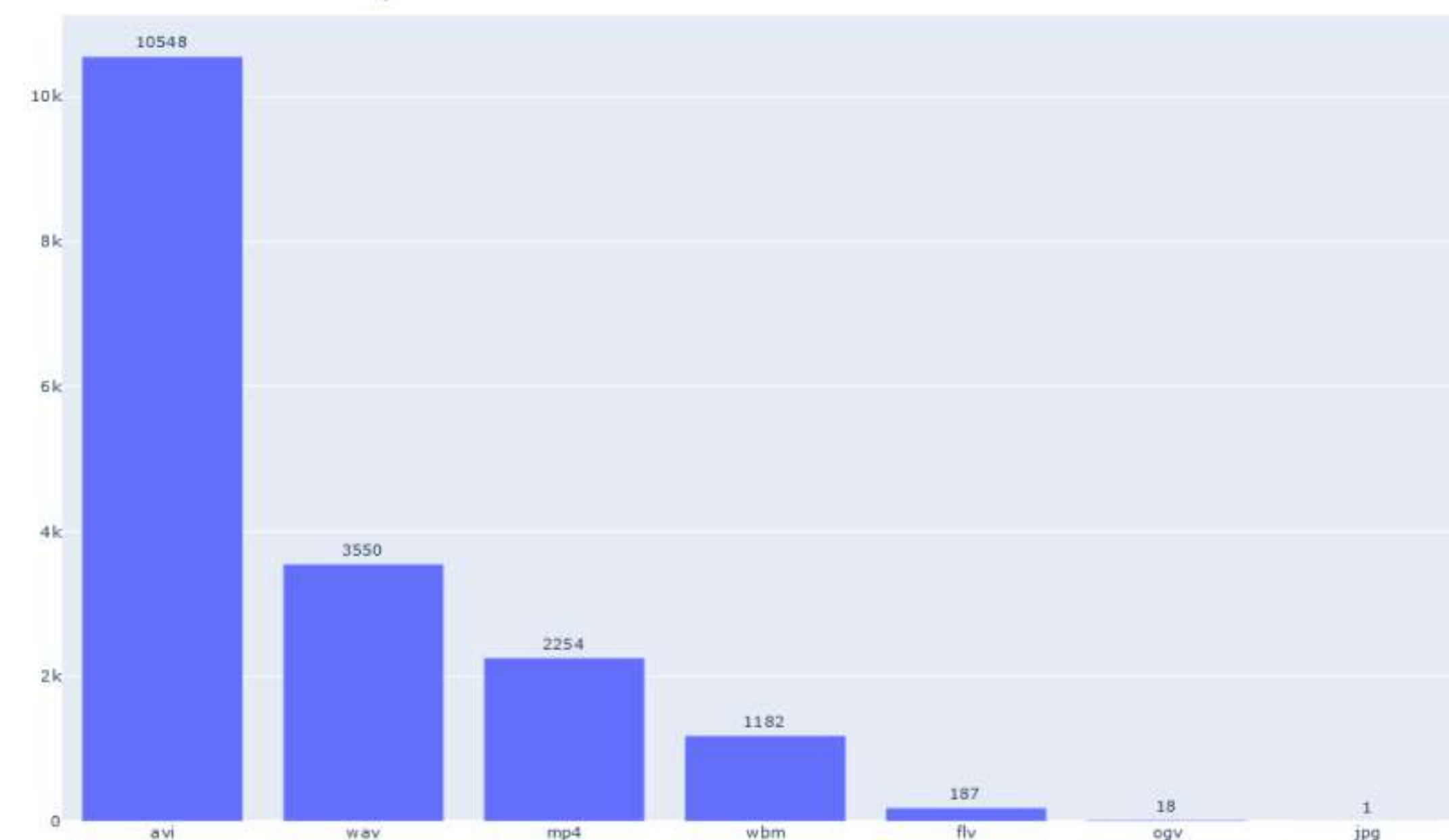
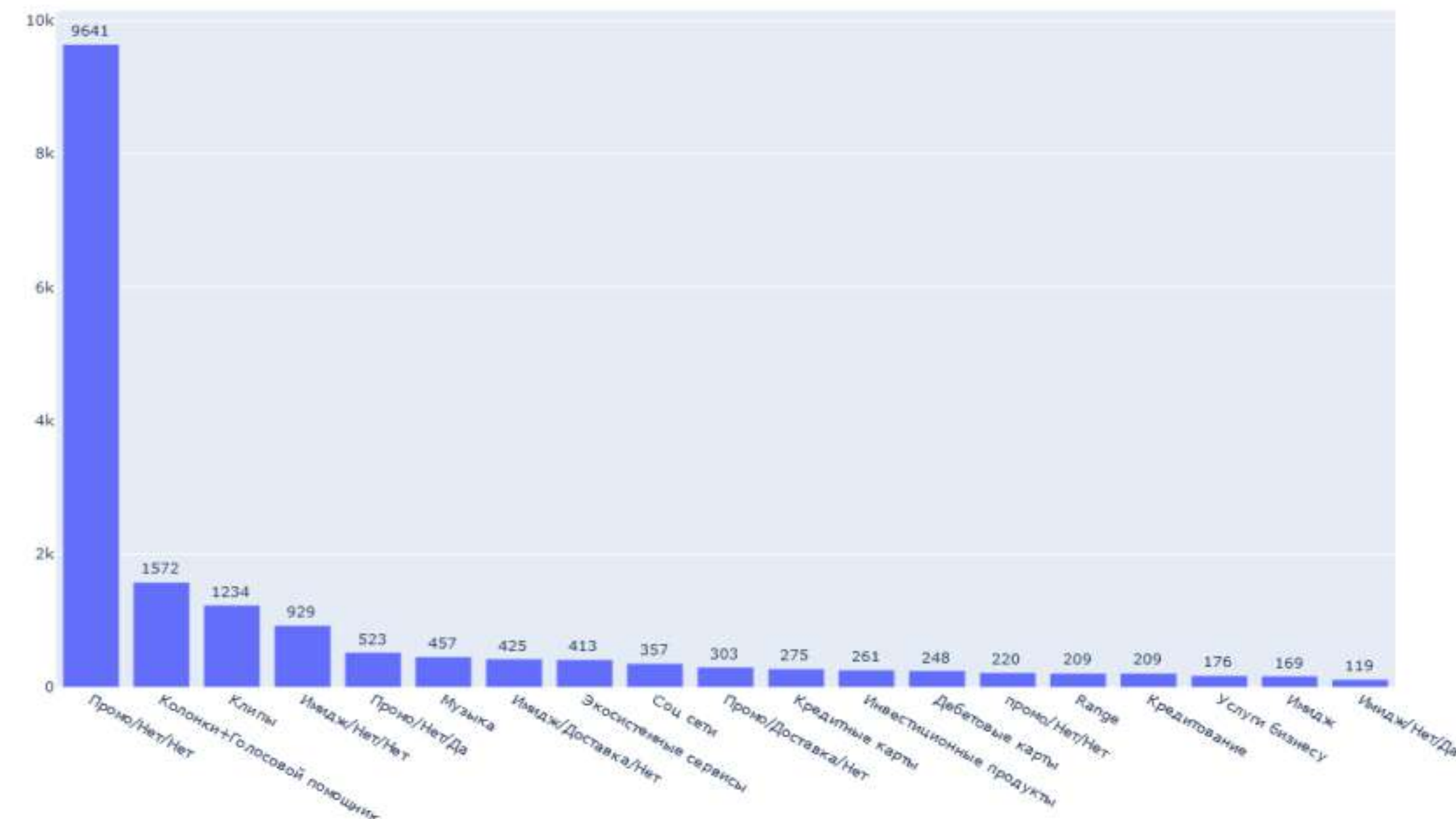
1. Высокая точность сегментации;
2. Масштабируемость решения
3. Адаптивность решения под меняющиеся реалии рынка
4. Высокая скорость работы

Know Your Data

- **17740** Видео и аудиофайлов (а также одинокий .jpg)
- **19(18)** классов. Их число может меняться со временем
- Сильный **дисбаланс** классов

Идея

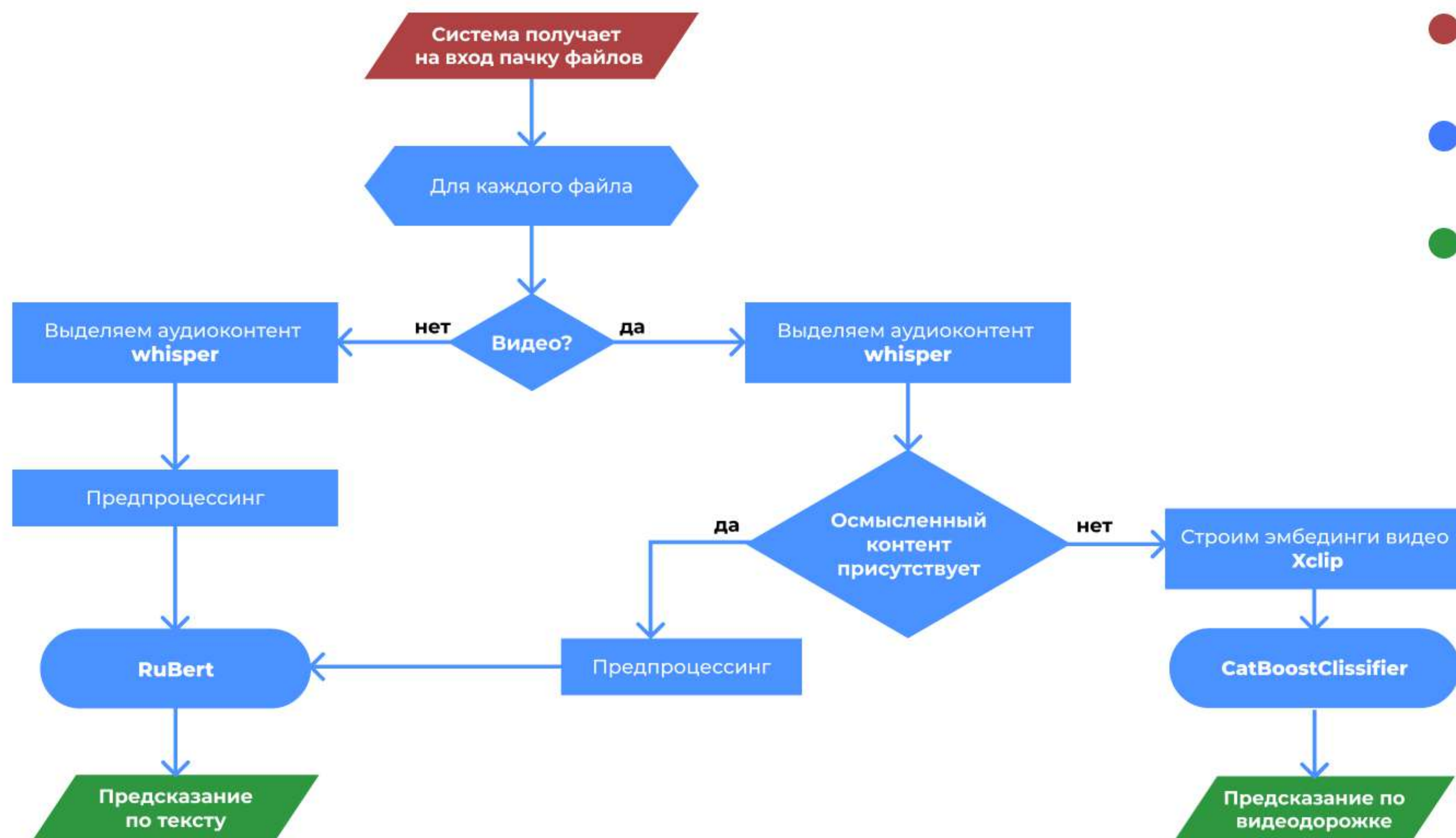
Потенциальное решение (скорее всего) лежит в **Open-Source** и должно хорошо работать «**из коробки**».



Наше решение

Как это работает

6



- начало
- основная работа алгоритма
- получаем на выходе

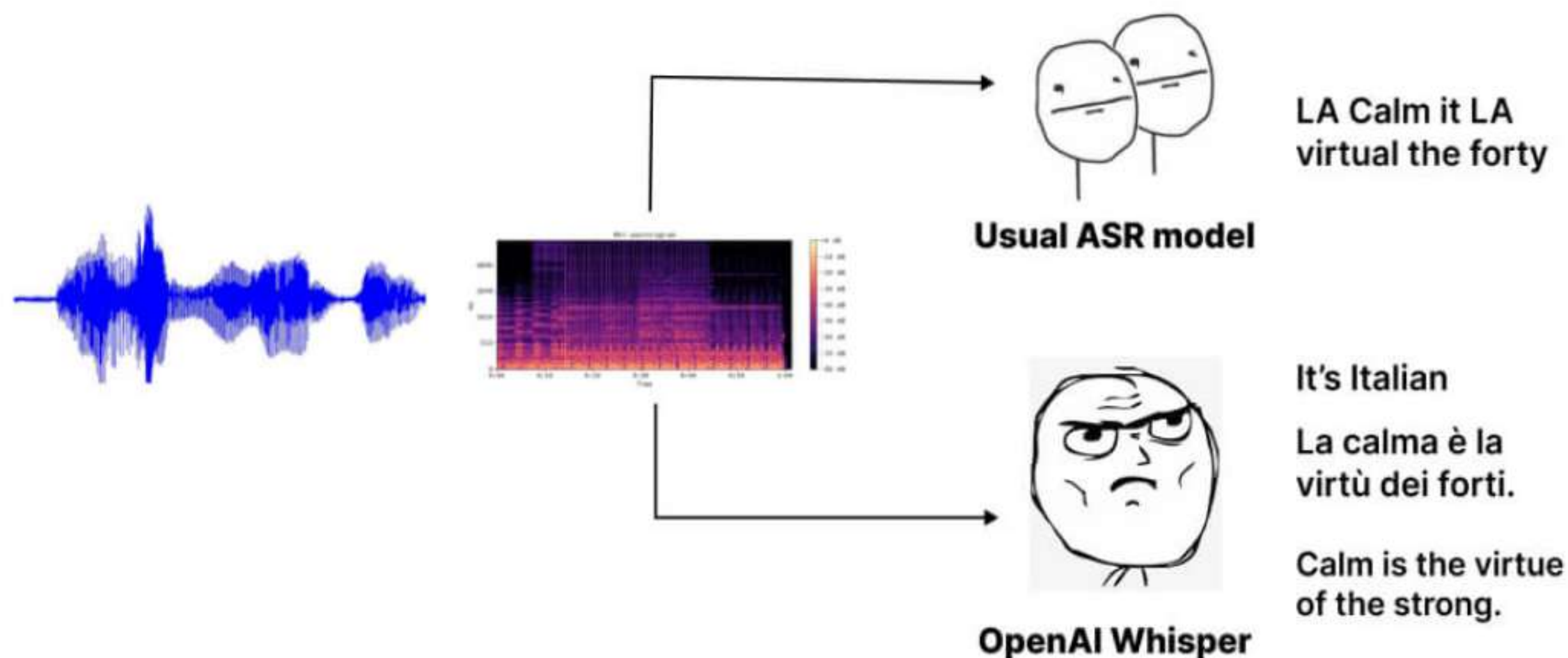


Ссылка на решение

Speech Recognition: whisper

7

Базовая гипотеза: в аудиодорожке к рекламе содержится важная для клиента (а значит и для нас) информация.



Решение: транскрибируем информацию из аудиофайла или звуковой дорожки видеоролика и проведем классификацию, используя извлеченную информацию.

Text classification: ruBERT

8

На основе извлеченных текстовых фрагментов была обучена модель семейства BERT, специализирующаяся на работе с русскоязычным текстом.

	precision	recall	f1-score	support
0	0.77	0.94	0.85	18
1	0.61	0.61	0.61	18
2	0.45	0.28	0.34	18
3	0.69	0.61	0.65	18
4	0.78	0.78	0.78	18
5	0.62	0.72	0.67	18
7	0.88	0.41	0.56	17
8	0.82	0.82	0.82	17
9	0.85	0.61	0.71	18
10	0.54	0.76	0.63	17
11	0.79	0.83	0.81	18
12	0.93	0.76	0.84	17
13	0.90	1.00	0.95	18
14	0.74	0.93	0.82	15
15	0.75	0.33	0.46	9
16	0.81	1.00	0.89	17
17	0.83	0.77	0.80	13
18	0.59	0.87	0.70	15
accuracy			0.73	299
macro avg	0.74	0.73	0.72	299
weighted avg	0.74	0.73	0.72	299

Привет! Это снова я, Борис Петров. На Восьмое марта я сделал подарки всем девушкам в отделе. Маше, Оле..... А мне это было совсем нетрудно. Ведь с кредитной картой «ИМЯ БАНКА» я могу купить подарки прямо сейчас, а оплатить потом.....

```
prediction = classifier.predict(text)
print(f"Предсказанный класс: {dd[prediction]}")
```

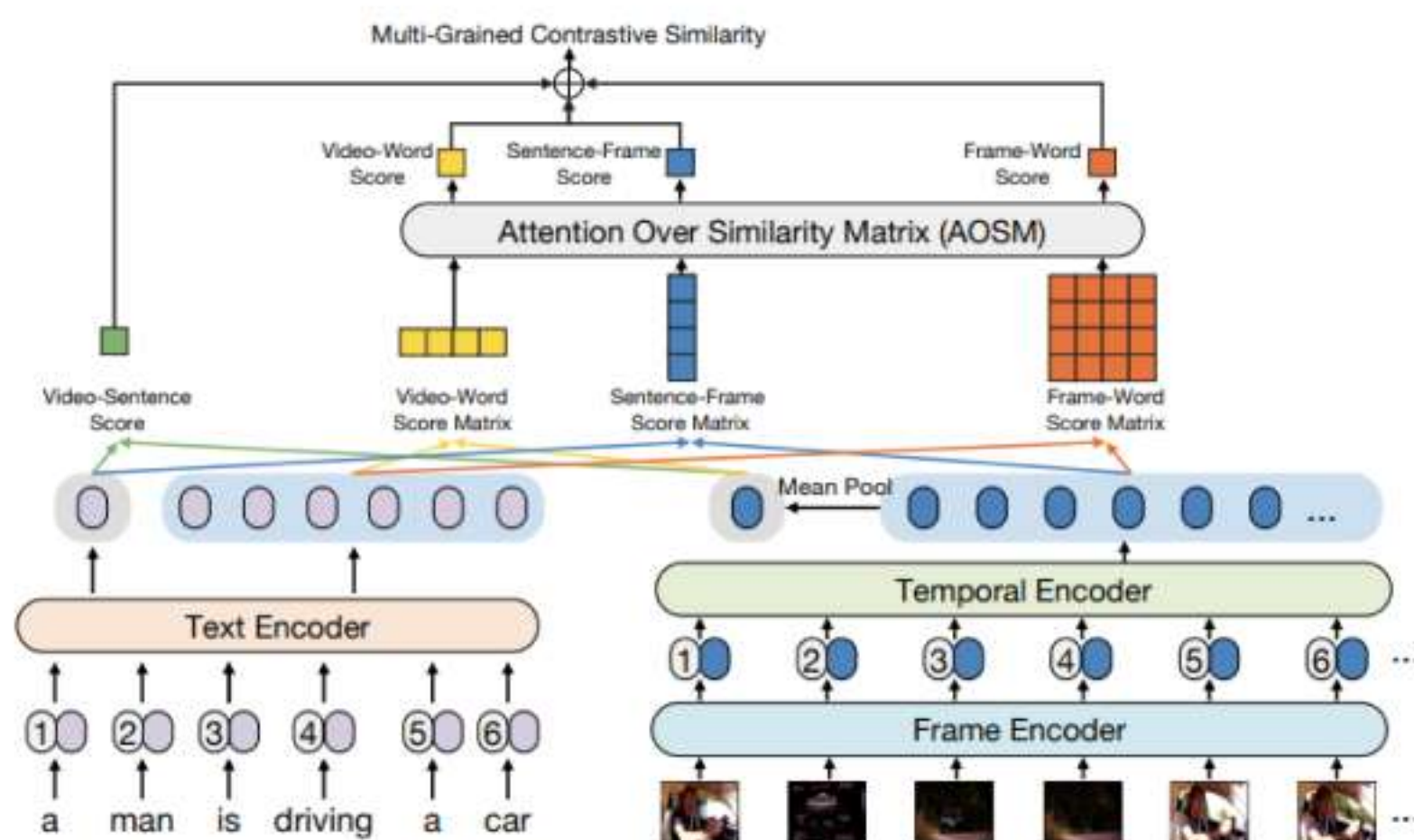
```
<
Предсказанный класс: Кредитные карты
CPU times: user 121 ms, sys: 0 ns, total: 121 ms
Wall time: 63.4 ms
```


Что делать если мы не смогли извлечь текст?

Гипотеза: В рекламных материалах, в которых присутствует только музыка и нет закадрового голоса, основной способ подачи материала (очевидно) визуальный. Мы используем два подхода для работы с файлами такого типа

Подход #1: StraightForward

Использовать видеоматериалы и эмбединги построенные на их основе для классификации



Подход #2: Generative AI

Применить генеративные сети, для генерации текстового описания видеофайла для последующей классификации



Video embeddings: X-clip+CatBoost

	precision	recall	f1-score	support
0	0.91	0.76	0.83	171
1	0.23	0.23	0.23	13
2	0.20	0.50	0.29	2
3	0.50	0.50	0.50	6
4	0.47	0.58	0.52	12
5	0.14	0.33	0.20	9
7	0.38	0.75	0.50	4
8	0.11	0.33	0.17	3
9	0.00	0.00	0.00	4
10	0.50	0.40	0.44	5
11	0.00	0.00	0.00	2
12	0.50	0.40	0.44	5
13	0.50	0.50	0.50	4
14	0.40	0.75	0.52	8
15	0.35	0.60	0.44	10
16	0.82	0.66	0.73	35
17	0.81	0.50	0.62	26
18	0.33	0.50	0.40	6
accuracy			0.64	325
macro avg	0.40	0.46	0.41	325
weighted avg	0.73	0.64	0.67	325

Получаем внутреннее представление видео (эмбединг) от X-clip и используем полученные эмбединги для классификации



CatBoost

Однако, к сожалению, результаты применения такого подхода далеки от идеальных и требуют, вероятно, больше времени для совершенствования

Describe the AD is provided. What is the product?'

The product is a credit card, as shown in the video.</s>

Кредитные карты

5.6870152950286865

The video does not provide any information about the product. However, the video shows a blue and white logo with the words "Shahzad Music Production" and "HMS Codec" on a blue background.</s>

Музыка

7.824151515960693

The product being advertised is a music app called YouTube Music. The video showcases the app's features and how it can be used to listen to music.</s>

Музыка

7.085063219070435

What is the product in the AD? I think it may belongs to following categories: ...

The product in the AD is a music app called "YouTube Music".</s>

Музыка

5.804338216781616

The product in the AD is a wireless headset.</s>

Колонки+Голосовой помощник

5.849937677383423

The product in the AD is a music video.</s>

Музыка

5.591860771179199

Mempuku

BERT

XCLIP
+
COSINE
SIMILARITY

XCLIP
+
CATBOOST

f1score_macro:
0.72

f1score_macro:
0.42

f1score_macro:
0.41

Total Inference: ~2 sec

Dashboards + Inference: Streamlint + FastAPI

Загрузите csv файл

Drag and drop file here
Limit 200MB per file

Browse files

обработать csv

Загрузите видео и аудио файлы

Drag and drop files here
Limit 200MB per file

Browse files

4447674.avi 4.2MB

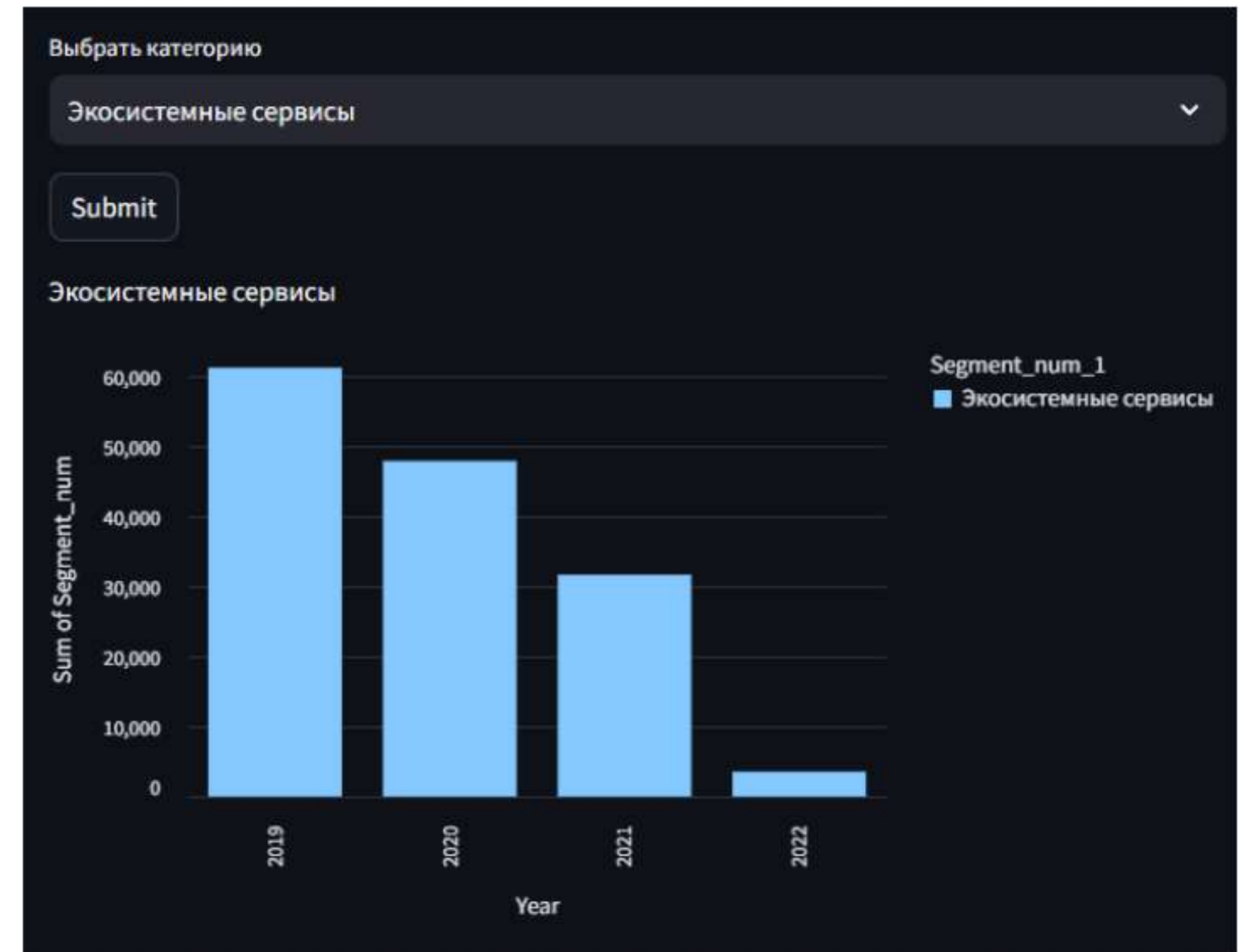
url видео

обработать файлы

	File	Category
0	4447674.avi	Услуги бизнесу

Download data as Excel

- Время инференса на одно видео ~ 2с.
- Загрузка и выгрузка Excel файла
- Загрузка видео в ручном режиме
- Загрузка видео с популярных видеохостингов
- Работа в интерфейсе и по API



Построение дашбордов согласно техническому заданию

Roadmap

14



Переразметка и насыщение

Generative AI

AutoML + Automatization + ...

Application support

Спасибо **за**
ваше **внимание**

