

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение

высшего образования

«Южно-Уральский государственный университет  
(национальный исследовательский университет)»

Высшая школа электроники и компьютерных наук

Кафедра системного программирования

# **Предобработка аудио с извлечением метаданных для обучения нейросетевой модели по рекомендации музыки**

Рецензент:

Руководитель проекта

ООО «Максимал»

Н.В. Баргая

Автор:

Студент группы КЭ-229

И.А. Березин

Научный руководитель:

доцент кафедры СП, к.ф.-м.н., доцент

И.И. Клебанов

# АКТУАЛЬНОСТЬ

1. Эпоха развития нейросетевых технологий, технологии машинного обучения и искусственного интеллекта (ИИ).
2. Большинство современных музыкальных сервисов работают по устаревшим технологиям.
3. Хранение и передача аудио и музыкальных данных занимает длительное время, с помощью нейросетевых технологий, можно значительно сократить этот показатель.

# Цели и задачи работы

## Цель:

Целью работы является предобработка аудио с извлечением метаданных для обучения нейросетевой модели по рекомендации музыки, благодаря этой системе можно будет запускать генератор наборов данных, для обучения модели на музыке по различной тематике.

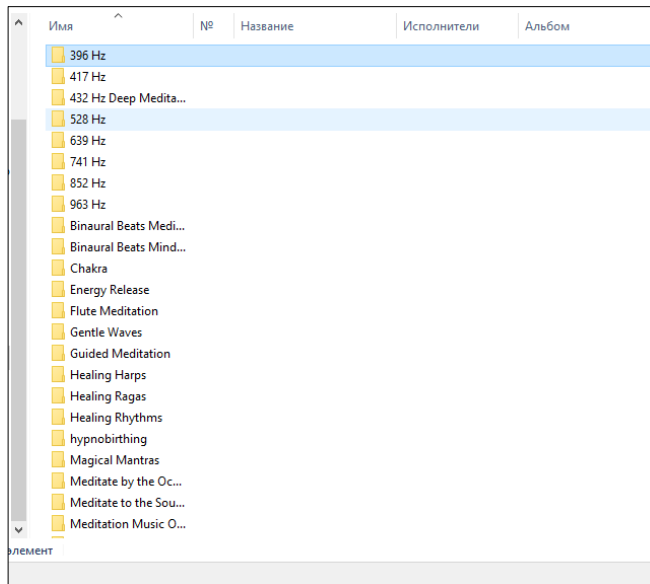
## Задачи:

1. Проанализировать предметную область.
2. Определить структуру и необходимые преобразования данных.
3. Реализовать генератор обучающего набора данных.
4. Наладить автоматизацию с базой данных и возможность работы с файлами на сервере.
5. Обучить модель.

# Обработка исходных данных.

## Требования к системе

1. Система должна распределять пары аудиофайл-метаданные по алгоритму нормального распределения
2. Система должна проверять пары аудиофайл-метаданные на допустимость их использования при создании наборов данных
3. Система должна иметь возможность взаимодействовать с парами аудиофайл-метаданные локально и удаленно.
4. Система должна иметь возможность хранить пары аудиофайл-метаданные в БД.



Имя	№	Название	Исполнители	Альбом
396 Hz				
417 Hz				
432 Hz Deep Medita...				
528 Hz				
639 Hz				
741 Hz				
852 Hz				
963 Hz				
Binaural Beats Medi...				
Binaural Beats Mind...				
Chakra				
Energy Release				
Flute Meditation				
Gentle Waves				
Guided Meditation				
Healing Harps				
Healing Ragas				
Healing Rhythms				
hypnobirthing				
Magical Mantras				
Meditate by the Oc...				
Meditate to the Sou...				
Meditation Music O...				



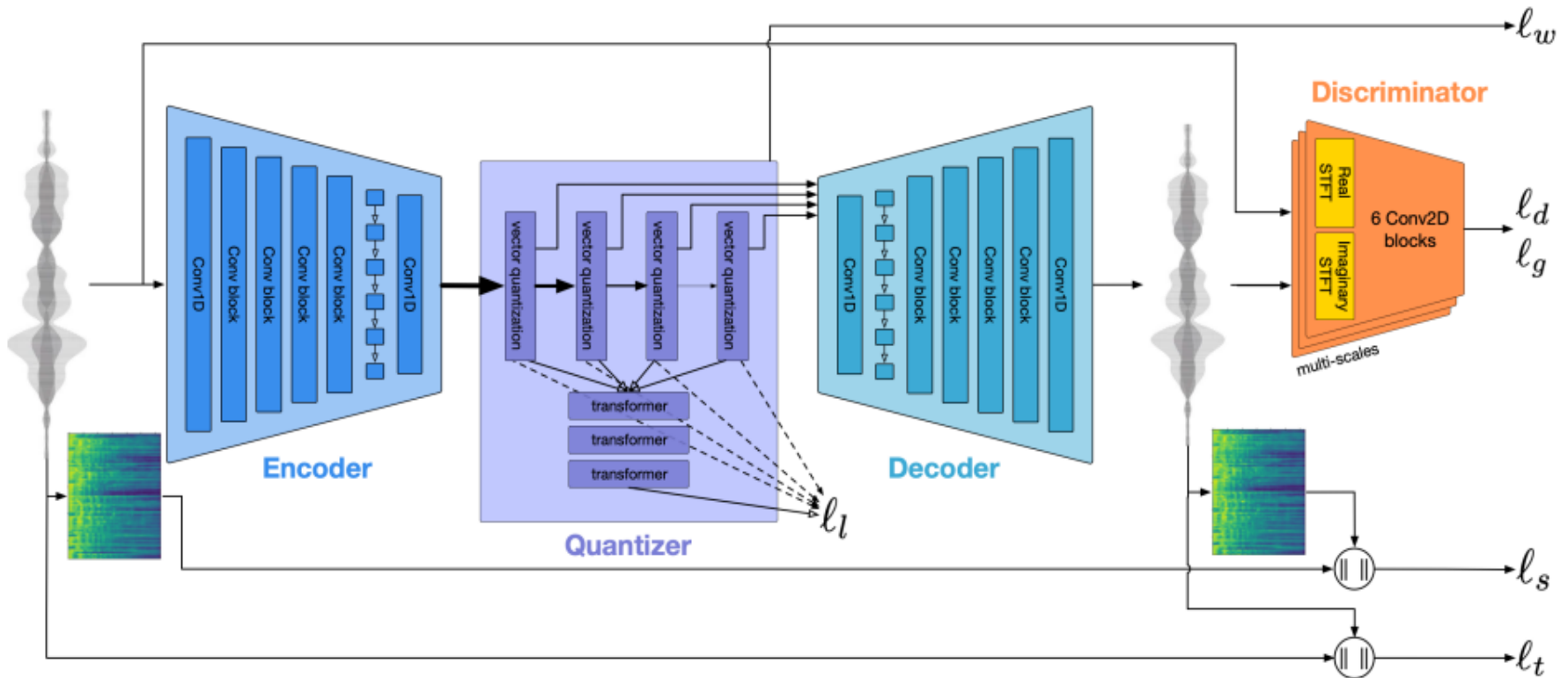
Имя	№	Название	Исполнители	Альбом
01 8 Chakras Bowls...	1	8 Chakras Bowls	Mudra Namaste	Salutation Bowls
01 963 Hz El Futuro...	1	963 Hz El Futuro	Paz Interna	963 Hz El Futuro
01 963 Solfeggio Att...	1	963 Solfeggio Attunement	Healing Solfeggio ...	963 Path to Awakening
01 963 Solfeggio.mp3	1	963 Solfeggio	Melatonement	963 Solfeggio
01 963hz Connect t...	1	963hz Connect to Source ...	Solfeggio Frequen...	963hz Connect to So...
01 Alpha bowls 10h...	1	Alpha bowls 10hz	Dream Wizard	Alpha bowls 10hz
01 Awakenings 963 ...	1	Awakenings 963 Hz	Hertz-Hunters	Awakenings 936 & 43...
01 Chakra Healing....	1	Chakra Healing	Makalu	Chakra Healing
01 Clear Thoughts 9...	1	Clear Thoughts 963 Hz	Ozonezzz	Clear Thoughts 963 Hz
01 Crown Chakra - ...	1	Crown Chakra - Asian Flu...	Massage Tribe	Chakra Healing - Ma...
01 Divine Hemisphe...	1	Divine Hemispheres 963 Hz	Zephyrical	Divine Hemispheres ...
01 Golden Intuition ...	1	Golden Intuition 963Hz	Introspective Rele...	Golden Intuition 963Hz
01 Letting Go.mp3	1	Letting Go	Frozen Voices	Letting Go
01 Miracles 963 Hz...	1	Miracles 963 Hz	Drone-Dax	Miracles 963 Hz
01 Nagnallar.mp3	1	Nagnallar	Sahasrara Beats	Nagnallar
01 Orange Waves 9...	1	Orange Waves 963 hz	Afar Lux	Orange waves
01 Point Cabrillo.mp3	1	Point Cabrillo	Lundegard	Point Cabrillo
01 Solar Waves.mp3	1	Solar Waves	Tejal Yann	Solar Waves
01 The Sun Rises 96...	1	The Sun Rises 963 Hz	Tina Cirdan	The Sun Rises 963 Hz
01 Unrushed.mp3	1	Unrushed	Solfeggio Dreams	69,3 Hz Atmospheres...
01 Vast and Blue.mp3	1	Vast and Blue	Gloaming Days	Vast and Blue
01 Vi%U0301%U00F...	1	Viðsyn	Traum Fanger	Viðsyn
02 Chakra Healing ...	2	Chakra Healing Bowl (Sou...	EXOPIRS	Meditation Sounds

# «AudioCraft» от Facebook \*

1. «MusicGen» - алгоритм генерации музыки построенный на одноступенчатой модели LM, которая работает с сжатыми дискретными музыкальными токенами.
2. «AudioGen» - авторегрессивная генеративная модель, которая создает аудиообразцы на основе текстовых входных данных аналогично «MusicGen» .
3. «EnCodec» - современный аудио кодек, обеспечивающий высокую точность и работающий в реальном времени за счет использования нейронных сетей.

\* Компания Facebook (старое название), ассоциируется с современным названием Meta – данная организация запрещена на территории РФ.

# Архитектура «EnCodec»



# Используемые инструменты и технологии

1. Среда разработки – PyCharm.
2. Основной язык программирования Python 3.10.4
3. База данных – MongoDB (класс NoSQL). \*
4. Библиотеки для работы с аудио – librosa, audio-metadata, eyed3, ffprobe (компонент ffmpeg).
5. Фреймворк – Hydra, для создание конфигурации системы.
6. Контроль разработки – GitLab.
7. Безопасность, серверная часть – библиотека paramiko, технологии SSh и SFTP с RSA ключами.\*

\* Дополнительные компоненты системы

# Структура метаданных

audiocraft / dataset / example / electro\_1.json 



JadeCopet AudioCraft v1.0.0 release with training code, AudioGen, MultiBandDiff...



Code

Blame

1 lines (1 loc) · 322 Bytes

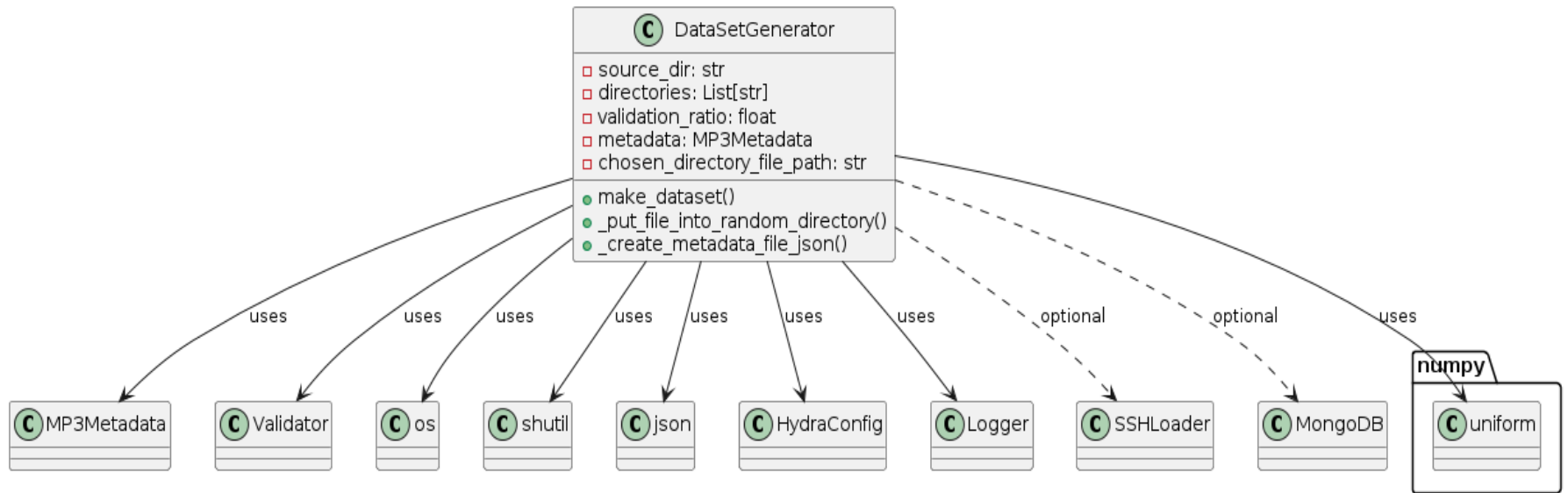
```
1 {"key": "", "artist": "Voyager I", "sample_rate": 48000, "file_extension": "mp3", "description":
```

- |   |                                    |
|---|------------------------------------|
| 1) «key» – ключ (тональность);          | 6) «keywords» – ключевые слова;    |
| 2) «artist» – исполнитель;              | 7) «duration» – продолжительность; |
| 3) «sample_rate» – частота;             | 8) «bpm» – бит;                    |
| 4) «file_extension» – расширение файла; | 9) «genre» – жанр;                 |
| 5) «description» – текстовое описание;  | 10) «title» – заголовок;           |
|   | 11) «name» – название;             |
|   | 12) «instrument» – инструменты;    |

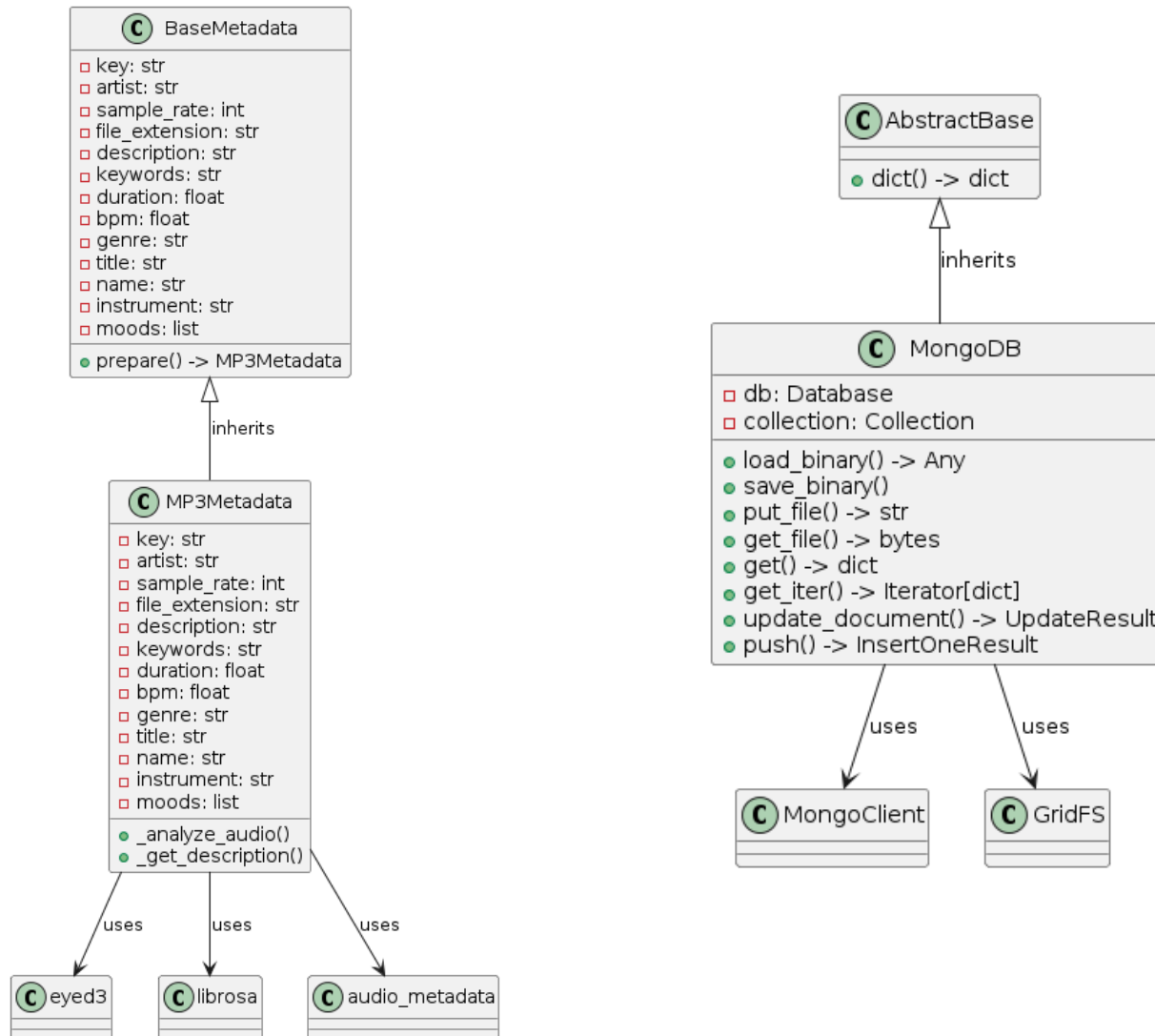


# Проектирование.

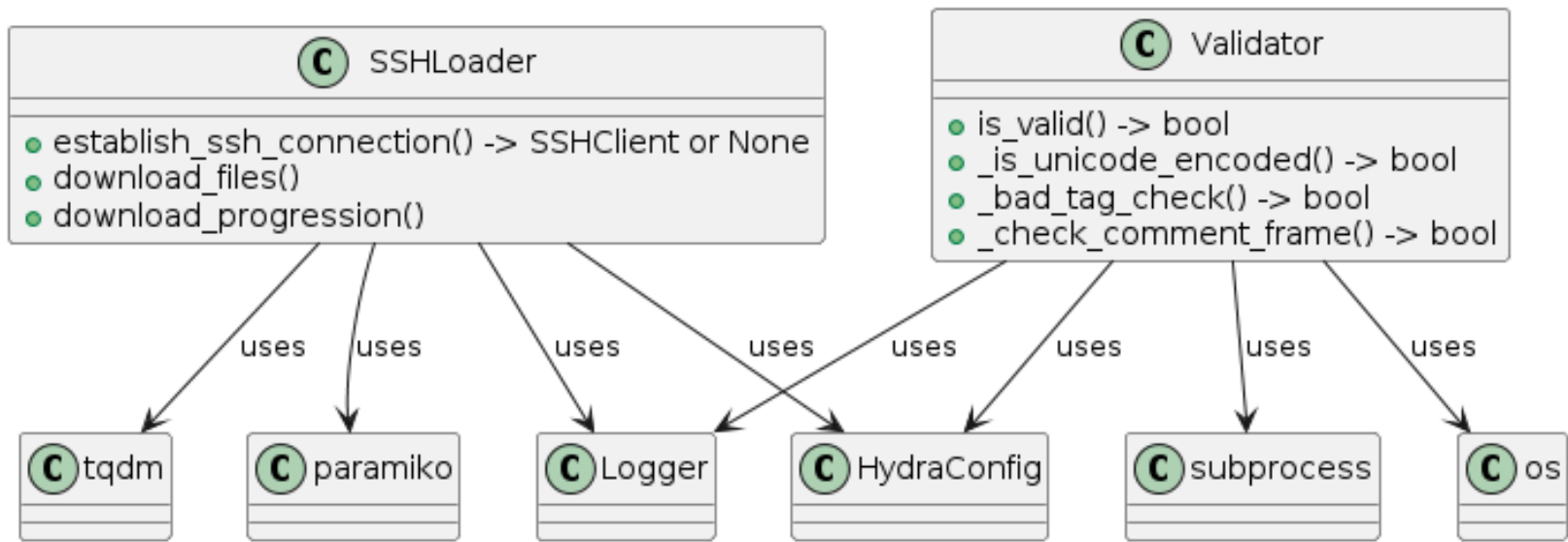
## Диаграмма КОМПОНЕНТОВ



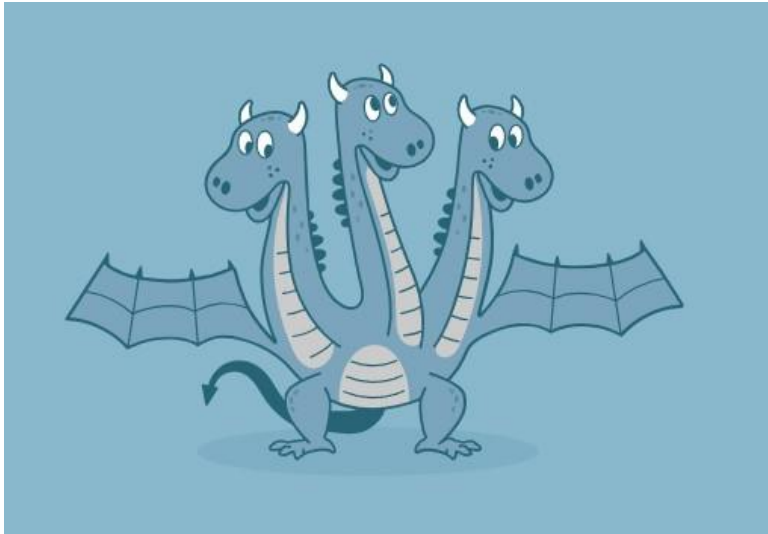
# Диаграмма КОМПОНЕНТОВ «MP3Metadata» и «MongoDB»



# Диаграмма КОМПОНЕНТОВ «SSHLoader» и «Validator»



# Конфигурирование и логгирование



```
config.yaml x
1 #config.yaml
2 defaults:
3   - local/default
4   - ssh/default
5   - db/mongo/default
6   - _self_
7
8 validation_ratio: 0.3
```

1. DEBUG. Используется для отладочной информации.
2. INFO. Используется для информационных сообщений.
3. WARNING. Используется для сообщений предупреждений.
4. ERROR. Используется для сообщений об ошибках.
5. CRITICAL. Используется для сообщений о критических ошибках.

# Извлечение метаданных (bpm и key)

```
tonalities = [  
    "C", "C#", "D", "D#",  
    "E", "F", "F#", "G",  
    "G#", "A", "A#", "B"  
]
```



```
y, sr = librosa.load(mp3)
```

```
bpm = librosa.beat.beat_track(y=y, sr=sr)
```

```
tone = librosa.effects.feature.tonnetz(y=y, sr=sr)
```

```
index = tone.sum(axis=1).argmax()
```

```
Key = tonalities[index]
```

# Проверка файлов

1. Ошибки на этапе извлечения метаданных. Многие теги метаданных невозможно извлечь или отсутствие метаданных.
2. Ошибки на этапе обучения. «Incorrect BOM value. Error reading comment frame» - ошибка чтения тега комментариев с указанием других тегов.
3. Ошибка чтения файлов. Проверка открытия файла и соответствие кодировки «unicode».



**FFmpeg**

# Генерация набора данных

Функция «`numpy.random.uniform`» принимает на вход три значения.

1. Параметр «`low`» – нижняя граница интервала. Генерируемые числа будут больше или равны этому значению. Стандартное значение – 0.
2. Параметр «`high`» – верхняя граница интервала. Генерируемые числа будут меньше этого значения. Стандартное значение – 1.
3. Параметр «`size`» – форма возвращаемого массива. Если не указано, возвращается одно случайное число.

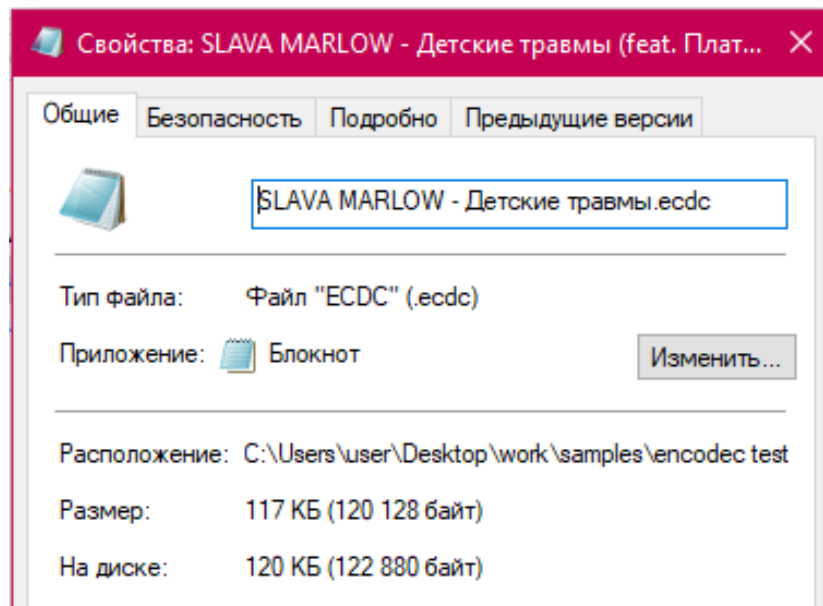
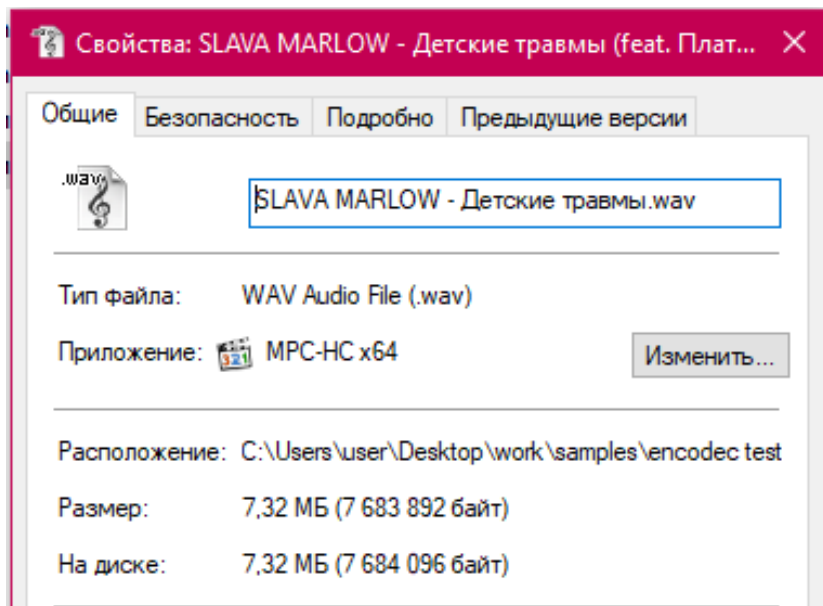


# Результат обучения модели «EnCodec»

```
[flashy.solver][INFO] - Train Summary [flashy.solver][INFO] - Train Summary
Epoch 1                               Epoch 90
bandwidth=2.200                         bandwidth=2.200
d_loss=1.923                             d_loss=1.395
d_msstftd=1.923                         d_msstftd=1.395
l1=0.053                                 l1=0.025
msspec=0.154                             msspec=0.052
penalty=0.423                            penalty=2.161
ratio1=0.444                             ratio1=0.617
g_loss=10.429                            g_loss=1.957
ratio2=132.702                           ratio2=22.191
mel=0.776                                 mel=0.362
sisnr=4.991                              sisnr=-6.549
adv=0.489                                 adv=0.804
feat=0.019                                feat=0.046
duration=31226.479                       duration=3014.021
```



# Токенизация (сжатие) с помощью модели «EnCodec»



Token example:

```
{ECDC ;{"m": "encodec_24khz", "al": 3841907, "nc": 8, "lm": true}; metadata: ...,  
vector: .....}
```

# Функциональное тестирование

№	Тест	Ожидаемый результат	Полученный результат	Прохождение теста
1	Тестирование взаимодействия с аудио файлами средствами ОС	Система читает, перемещает и удаляет аудио-файлы средствами ОС	Результат соответствует ожиданиям	Пройден
2	Тестирование извлечения метаданных с аудио	Система извлекает метаданные с аудио	Результат соответствует ожиданиям	Пройден
3	Тестирование распределения пар аудиофайл-метаданные по алгоритму нормального распределения	Система распределяет пары аудиофайл-метаданные по алгоритму нормального распределения	Результат соответствует ожиданиям	Пройден
4	Тестирование компонента «валидации», который проверяет файлы на допустимость их использования.	Компонент «валидации» успешно определяет плохие файлы и перемещает их в специальную директорию	Результат соответствует ожиданиям	Пройден
...	...	...	...	...
10	Тестирование установки необходимых пакетов из файла «requirements.txt»	Все пакеты актуальных версий успешно устанавливаются	Результат соответствует ожиданиям	Пройден

# Тестирование компонентов

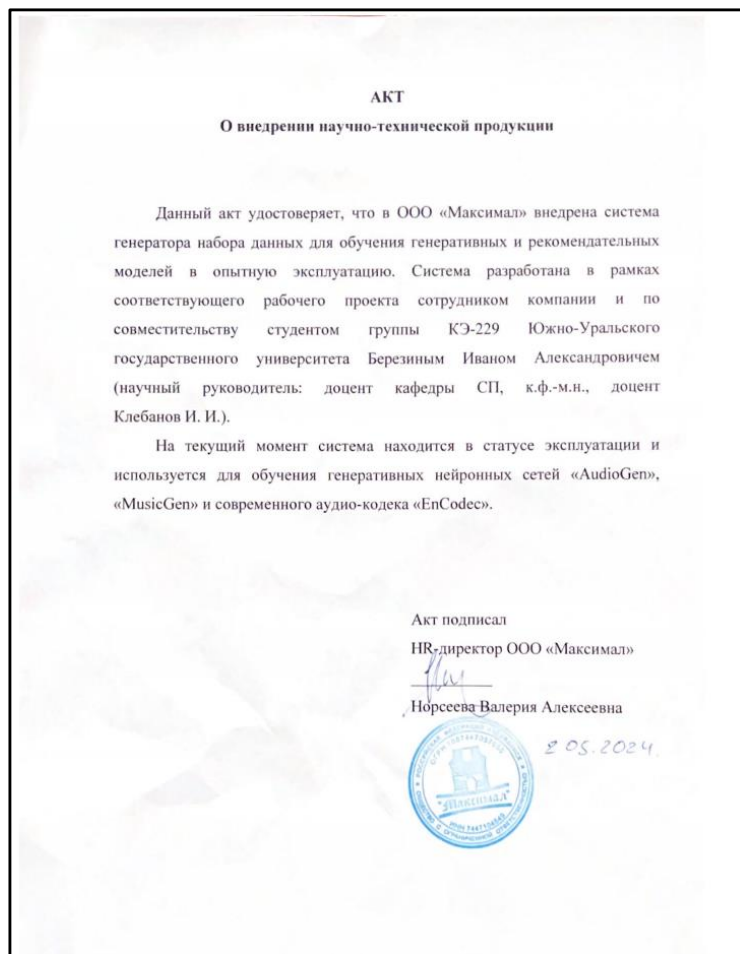
Для тестирования компонентов, не используемых во время работы основного модуля, были проведены отдельные тестовые сценарии с подключением соответствующих компонентов: «SShLoader» и «MongoDB».

Тестирование проводилось аналогично тестовым сценариям функционального тестирования.

Тесты пройдены успешно.

# Эксплуатация и дальнейшая разработка

На текущий момент система успешно эксплуатируется. Справка о введении в эксплуатацию предоставлена. Дальнейшая работа над системой включает расширение ее возможностей и вариативности использования с другими данными.



# Основные результаты

1. Проанализирована предметная область.
2. Определена структура и необходимые преобразования данных.
3. Налажена автоматизация с базой данных и работа с файлами на сервере.
4. Реализован генератор обучающего набора данных.
5. Проведено тестирование и обучение модели на обработанных данных.