МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Факультет прикладной математики, информатики и механики

Кафедра математических методов исследования операций

Отчет

о научно-исследовательской работе

«Анализ аудиоданных с помощью машинного обучения»

Направление 02.04.02 – Фундаментальная информатика и информационные технологии

Программа – Машинное обучение и интеллектуальные информационные технологии

Курс 1 группа 13 семестр 1

Обучающийся Лихачев М.Ю.

Руководитель к.ф.м.н Замятин И.В.

Воронеж – 2020

Содержание

Список сокращений 3

Введение 4

Глава 1. Аналитическая часть 6

1.1 Анализ предметной области 6

1.2 Анализ аудиоданных 7

1.2.1 Обзор аудиофайлов 7

1.2.2 Приложения по обработке звука 7

1.2.3 Аудиоданные и их признаки 7

1.3 Обзор существующих аналогов 14

1.3.1 Музыкальная платформа «Bandcamp» 14

1.3.2 Музыкальная платформа «Soundcloud» 15

1.3.3 Интернет-сервис потокового аудио «Spotify» 16

1.3.4 Сервис распознавания потокового аудио «Shazam» 17

1.3.5 Аудиоплеер «Apple Music» 18

1.3.6 Результат исследования существующих аналогов 19

Список использованных источников 21

Список сокращений

ВГУ − Воронежский Государственный университет;

WAV − Waveform Audio File;

WMA − Windows Media Audio;

MFCC − Мел-частотные кепстральные коэффициенты;

ПО − Программное обеспечение;

WWDC − Apple Worldwide Developers Conference.

Введение

Музыка — важная часть жизни любого человека, одно из древних видов искусства. В настоящее время она продолжает развиваться, в том числе при помощи современных технологий. В сети Интернет размещено большое количество онлайн-приложений и платформ, предназначенных для работы с оцифрованной звуковой информацией (музыкой). В данную категорию входят такие типы систем, как музыкальные проигрыватели (iTunes, Google Play Music, Яндекс.Музыка, Spotify), системы распознания и поиска музыки (Shazam), а также платформы для распространения пользовательской авторской музыки (SoundCloud, Bandcamp).

Аудиоанализ — область, включающая автоматическое распознавание речи (ASR), цифровую обработку сигналов, а также классификацию, тегирование и генерацию музыки — представляет собой развивающийся поддомен приложений глубокого обучения. Некоторые из самых популярных и распространенных систем машинного обучения, такие как виртуальные помощники Alexa, Siri и Google Home, — это продукты, созданные на основе моделей, извлекающих информацию из аудиосигналов.

Актуальность данного проекта подтверждается потребностью пользователей в сервисах, использующих машинный анализ аудиоданных.

Целью данной научно-исследовательской работы является разработка web-приложения «Аудио-хостинг авторской музыки», включающий в себя механизм анализа аудиоданных, который будет соответствовать современным требованиями пользователей подобных ресурсов.

Для выполнения данной цели были поставлены следующие *задачи*:

- анализ предметной области;
- обзор существующих аналогов;
- изучение аудиоданных;
- выбор, разработка, реализация и тренировка модели машинного обучения;
- формирование требований к функциональным модулям сервиса;
- выбор инструментария реализации;

- проектирование базы данных;
- разработка компонентов приложения;
- тестирование и отладка на сервере.

Объектом исследования является машинный анализ аудиоданных.

Предметом исследования является web-сервис «Аудио-хостинг авторской музыки» как интерфейс взаимодействия пользователя с механизмом машинного обучения.

В ходе данной научно-исследовательской работы были использованы следующие *методы исследования*:

- анализ научной литературы и других источников по изучаемой предметной области;
- системный анализ, изучение объекта исследования.

Основные положения, выносимые на защиту:

- модель машинного обучения анализа аудиоданных;
- структура базы данных;
- интерфейс и дизайн web-приложения;
- компоненты программы;
- работоспособность в условиях реального пользования данной системы.

Структура проекта. Данная научно-исследовательская работа состоит из содержания, перечня использованных сокращений, введения, 2 глав, заключения, списка использованной литературы, приложения и рисунков.

При написании научно-исследовательской работы были использованы методические указания. [1]

Глава 1. Аналитическая часть

1.1 Анализ предметной области

Музыка — искусство, средством воплощения художественных образов для которого являются звук и тишина, особым образом организованные во времени.

Предполагается, что появление современного человека произошло около 160 000 лет тому назад в Африке. Около 50 000 лет тому назад люди заселили все пригодные для жизни континенты. Поскольку все люди мира, включая наиболее изолированные племенные группы, обладают некоторыми формами музыки, историки пришли к выводу, что музыка должна была присутствовать у первых людей в Африке, до их расселения по планете. Предполагается, что после возникновения в Африке музыка существует по крайней мере 50 000 лет и постепенно превратилась в неотъемлемую часть человеческой жизни по всей планете. [2]

Как и любой вид искусства музыка развивается и тесно сплетается с другими областями человеческой деятельности (в том числе и с интернетом).

Распространение музыки в Интернете началось с появления формата MP3 (примерно с 1-го января 1999 года), сжимающего звуковые файлы до размеров, пригодных для передачи в Интернете при сохранении качества записи.

На сегодняшний день не представляется возможным, хотя бы примерно, подсчитать количество музыкальных сайтов и сайтов с музыкальным содержанием.

1.2 Анализ аудиоданных

1.2.1 Обзор аудиофайлов

Звуковые волны оцифровываются путем выборки из дискретных интервалов, известных как частота дискретизации (как правило, 44,1 кГц для аудио с CD-качеством, то есть 44 100 семплов в секунду).

Каждый семпл представляет собой амплитуду волны в определенном временном интервале, где глубина в битах (или динамический диапазон сигнала) определяет, насколько детализированным будет семпл (обычно 16 бит, т.е. семпл может варьироваться от 65 536 значений амплитуды).

В обработке сигналов семплинг — это преобразование непрерывного сигнала в серию дискретных значений. Частота дискретизации — это количество семплов за определенный фиксированный промежуток времени. Высокая частота дискретизации приводит к меньшей потере информации, но к большим вычислительным затратам. [3]

1.2.2 Приложения по обработке звука

- Индексирование музыкальных коллекций согласно их аудиопризнакам.
- Рекомендация музыки для радиоканалов.
- Поиск сходства для аудиофайлов (Shazam).
- Обработка и синтез речи — генерирование искусственного голоса для диалоговых агентов. [3]

1.2.3 Аудиоданные и их признаки

Звук представлен в форме аудиосигнала с такими параметрами, как частота, полоса пропускания, децибел и т.д. Типичный аудиосигнал можно выразить в качестве функции амплитуды и времени.

Некоторые устройства могут улавливать эти звуки и представлять их в машиночитаемом формате. Примеры этих форматов:

- wav (Waveform Audio File)
- mp3 (MPEG-1 Audio Layer 3)
- WMA (Windows Media Audio)

Процесс обработки звука включает извлечение акустических характеристик, относящихся к поставленной задаче, за которыми следуют схемы принятия решений, которые включают обнаружение, классификацию и объединение знаний.

Спектрограмма — это визуальный способ представления *уровня* или *громкости* сигнала во времени на различных частотах, присутствующих в форме волны. Обычно изображается в виде тепловой карты. Данные преобразуются в кратковременное преобразование Фурье. С его помощью можно определить амплитуду различных частот, воспроизводимых в данный момент времени аудиосигнала.[3]

Каждый аудиосигнал состоит из множества признаков.

К *акустическим признакам* относятся:

- *высота*, которая зависит от частоты колебаний: чем выше частота, тем выше звук;
- *сила*, которая зависит от амплитуды колебаний: чем больше амплитуда, тем сильнее звук;
- *длительность* (или долгота), которая связана с количеством колебаний во времени,
- *тембр* – индивидуальное качество его акустических признаков.

Колебания могут быть *периодическими и непериодическими*. В результате периодических колебаний возникают тоны, а в результате непериодических – шумы. Тоны имеют абсолютную высоту, а шумы – относительную.

В образовании звуков важна роль резонатора – замкнутой воздушной среды, где производится звук. Благодаря резонатору основной тон обогащается наслаивающимися на него обертонами — более высокими

тонами, число колебаний которых является кратным по отношению к числу колебаний основного тона. Это гармонические обертоны.

Тоны в резонаторе могут возникать и самостоятельно. Это *резонаторные* тоны. В этом случае резонаторы резонируют задней и передней частью отдельно. Удлинение и укорочение резонатора меняют его тоновую окраску. Область резонирования и его результат называются формантой.

Тембр звука – сложное явление, содержит основной тон, шум, гармонические обертоны и резонаторные тоны. [4]

Спектральные (частотные) признаки получаются путем преобразования временного сигнала в частотную область с помощью преобразования Фурье. К ним относятся частота основного тона, частотные компоненты, спектральный центроид, спектральный поток, спектральная плотность, спектральный спад и т.д. [3]

Спектральный центроид указывает, на какой частоте сосредоточена энергия спектра или, другими словами, указывает, где расположен “центр масс” для звука. Схож со средневзвешенным значением:

$$f_c = \frac{\sum_k S(K) f(k)}{\sum_k S(k)},$$

где $S(k)$ — спектральная величина элемента разрешения k, а $f(k)$ — частота элемента k.[3] Иллюстрация представлена в соответствии с рис. 1.1.

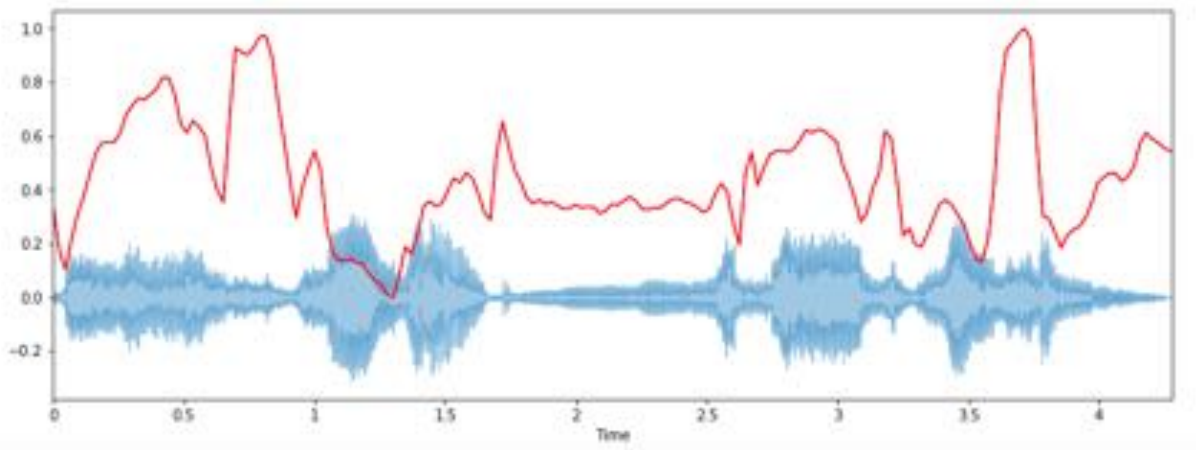

Рисунок 1.1 Спектральный центроид (иллюстрация)

Спектральный спад — это мера формы сигнала, представляющая собой частоту, в которой высокие частоты снижаются до 0. Чтобы получить ее, нужно рассчитать долю элементов в спектре мощности, где 85% ее мощности находится на более низких частотах.[3] Иллюстрация представлена в соответствии с рис. 1.2.

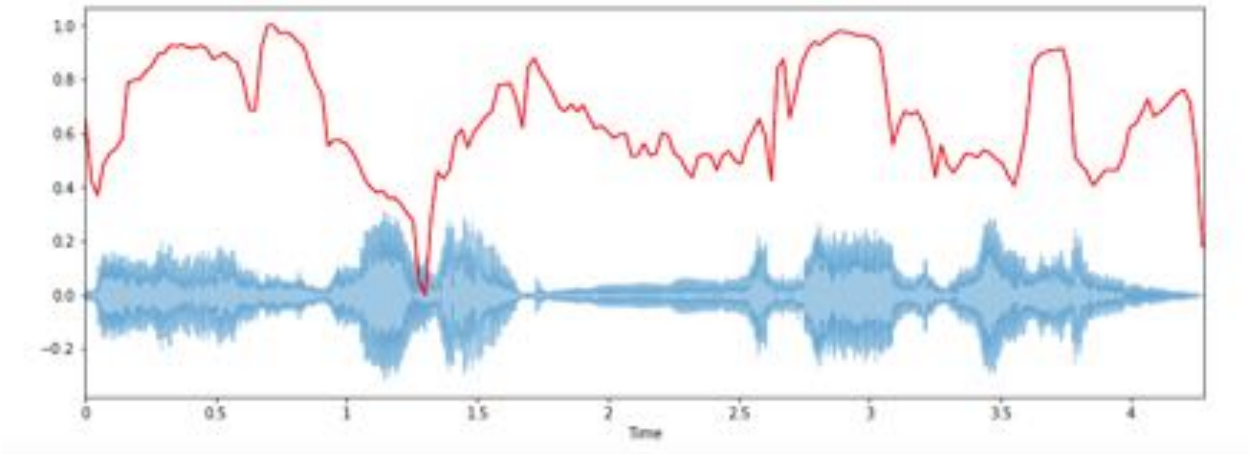

Рисунок 1.2 Спектральный спад (иллюстрация)

Спектральная ширина определяется как ширина полосы света на половине максимальной точки.[3] Иллюстрация представлена в соответствии с рис. 1.3.

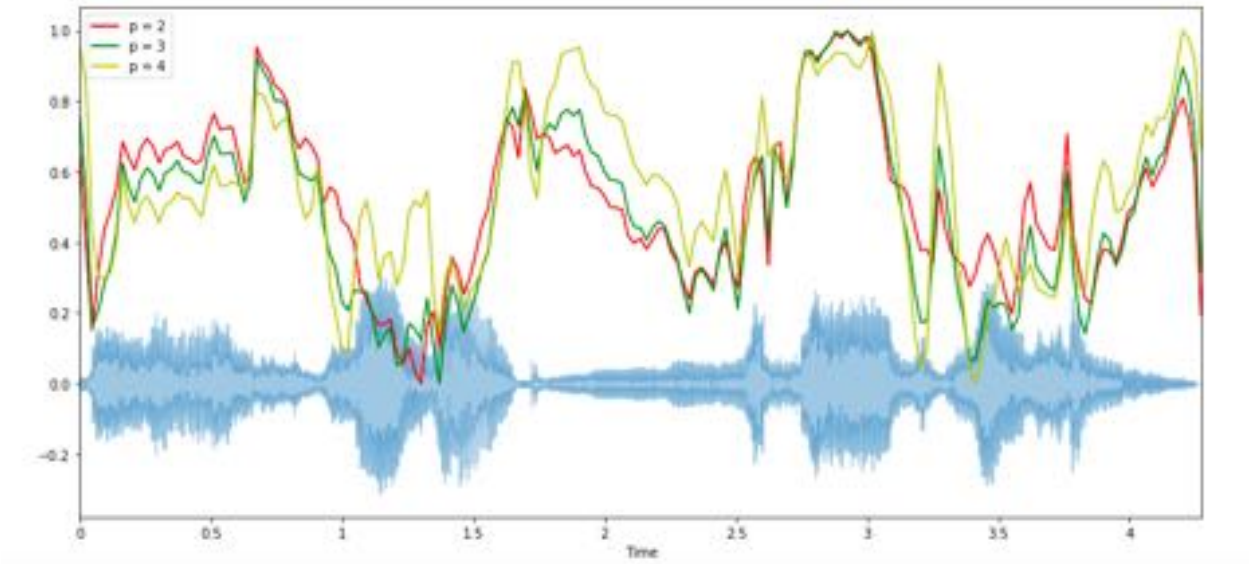

Рисунок 1.3 Спектральная ширина (иллюстрация)

Вычисление числа пересечений нуля в пределах сегмента этого сигнала — простой способ измерения гладкости сигнала. Голосовой сигнал колеблется медленно. Например, сигнал 100 Гц будет пересекать ноль 100 раз в секунду, тогда как “немой” фрикативный сигнал может иметь 3000 пересечений нуля в секунду.

$$zcr = \frac{1}{T-1}\sum_{t=1}^{T-1} II\{s_t - s_{t-1} < 0\},$$

где s_t — сигнал длиной t, а $II(X)$ — функция-индикатор (принимает значение 1, если выражение X принимает истинное значение, и 0 в противном случае).

Более высокие значения наблюдаются в таких высоко ударных звуках, как в металле и роке.[3] Иллюстрация представлена в соответствии с рис. 1.4.

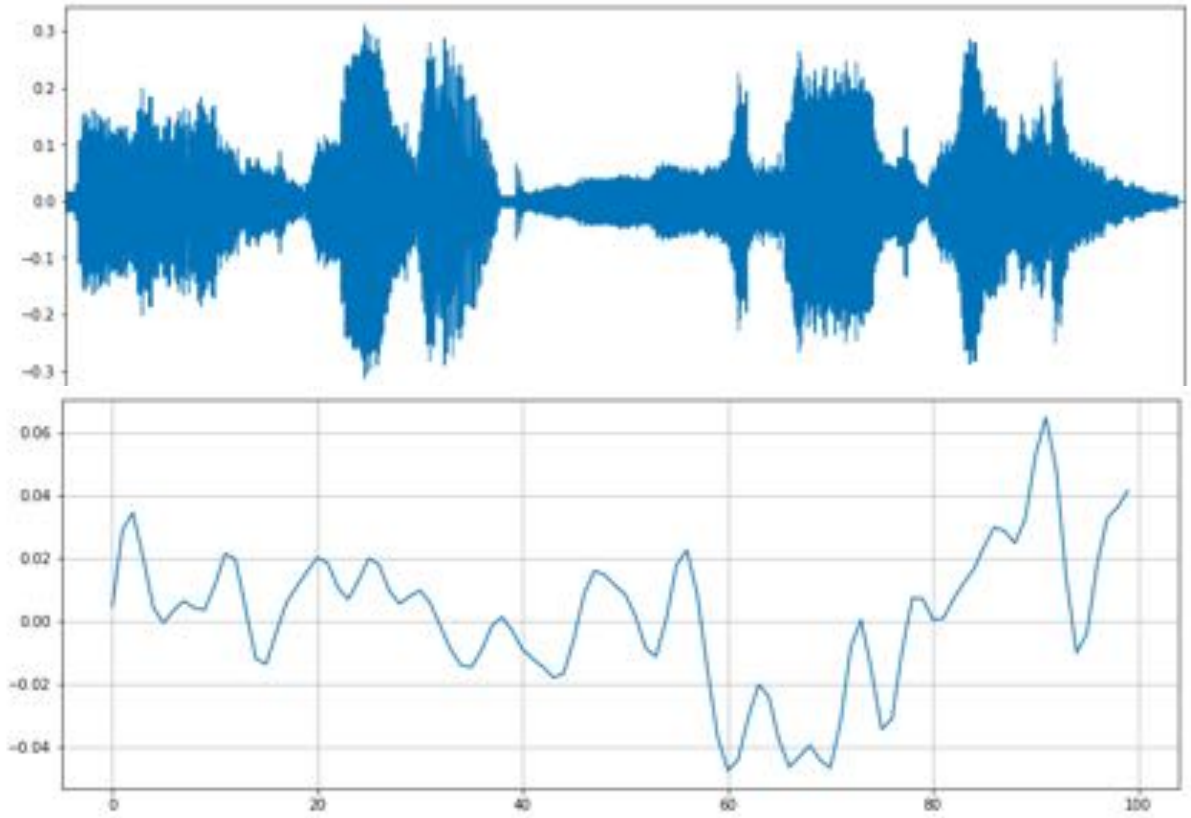

Рисунок 1.4 Пересечения нуля на примере аудиосигнала (иллюстрация)

Мел-частотные кепстральные коэффициенты (MFCC) - представляют собой небольшой набор признаков (обычно около 10–20), которые кратко описывают общую форму спектральной огибающей. Они моделируют характеристики человеческого голоса.[3] Иллюстрация представлена в соответствии с рис. 1.5.

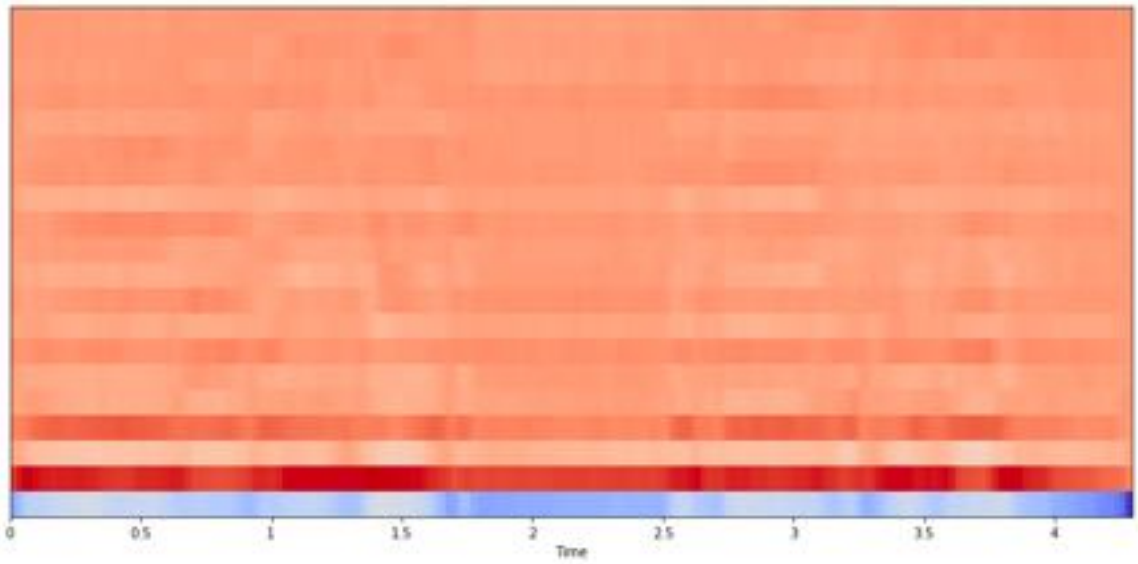

Рисунок 1.5 Отображение MFCC (иллюстрация)

Цветность (признак или вектор цветности) обычно представлен вектором признаков из 12 элементов, в котором указано количество энергии каждого высотного класса {C, C#, D, D#, E, F, F#, G, G#, A, A#, B} в сигнале. Используется для описания меры сходства между музыкальными произведениями.[3] Иллюстрация представлена в соответствии с рис. 1.6.

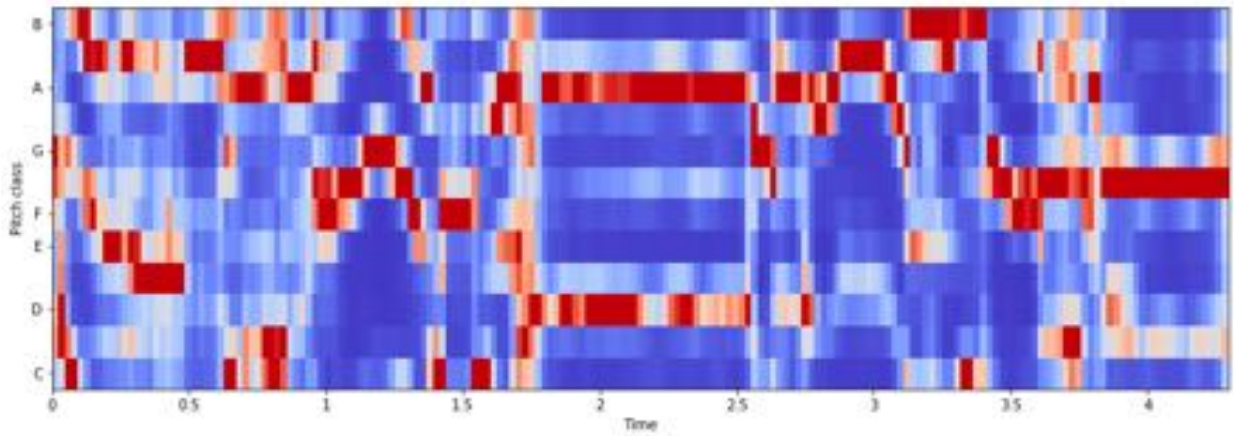

Рисунок 1.6 Признаки цветности (иллюстрация)

1.3 Обзор существующих аналогов

В ходе данной научно-исследовательской работы в качестве примеров были изучены и рассмотрены различные аудиосервисы. Данное ПО были отобраны как самые популярные среди опрошенной пользовательской аудитории.

1.3.1 Музыкальная платформа «Bandcamp»

Рассмотрим самый популярный из выбранных портал для распространения музыки. Вот его главная страница в соответствии с рисунком 1.7

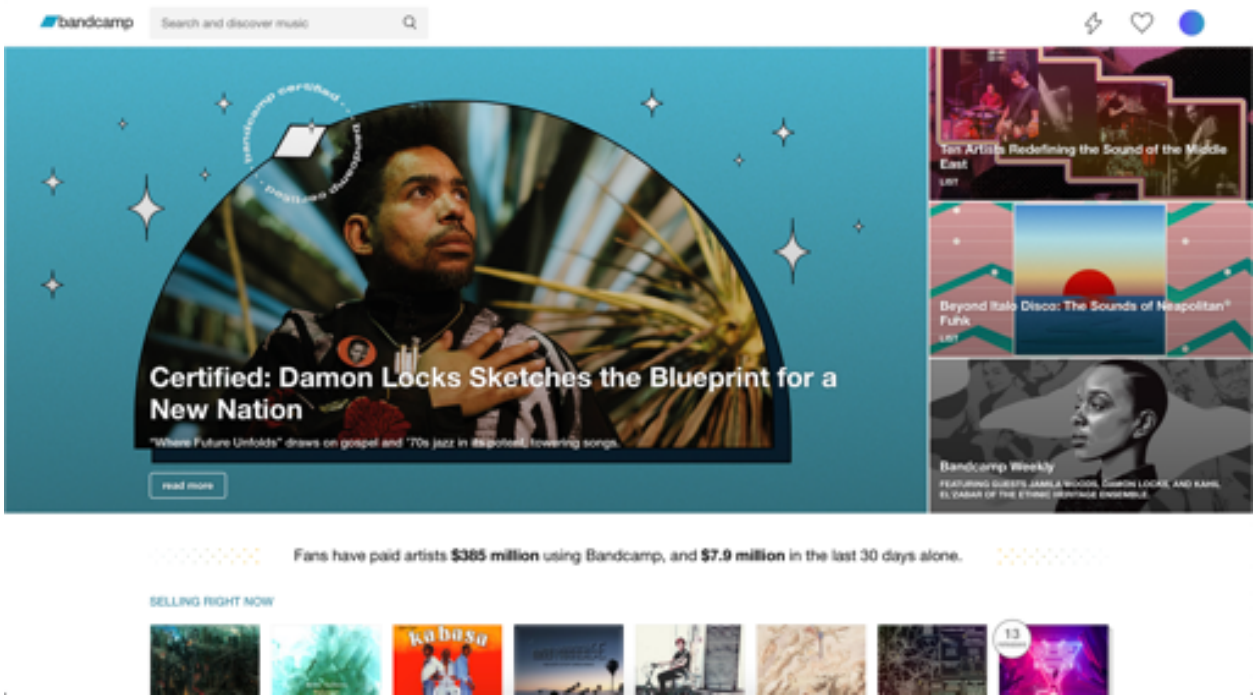

Рисунок 1.7 Главная страница «Bandcamp» (верхняя часть)

Сайт имеет довольно приятный на глаз современный дизайн в стиле минимализм. Предоставляет возможность пользователям делиться своей музыкой, в том числе и для коммерческих целей, и подписываться на других исполнителей.

При всем своем большом наборе положительных сторон, эта платформа имеет и ряд недостатков. Например, акцент приложения сделан больше на выкладывание собственного творчества, или подпиской на других исполнителей. Пример страницы музыкальной композиции представлен в соответствии с рисунком 1.8. Однако для начинающих авторов было бы очень полезно получить развернутую оценку от слушателей для более качественной работы в будущем.

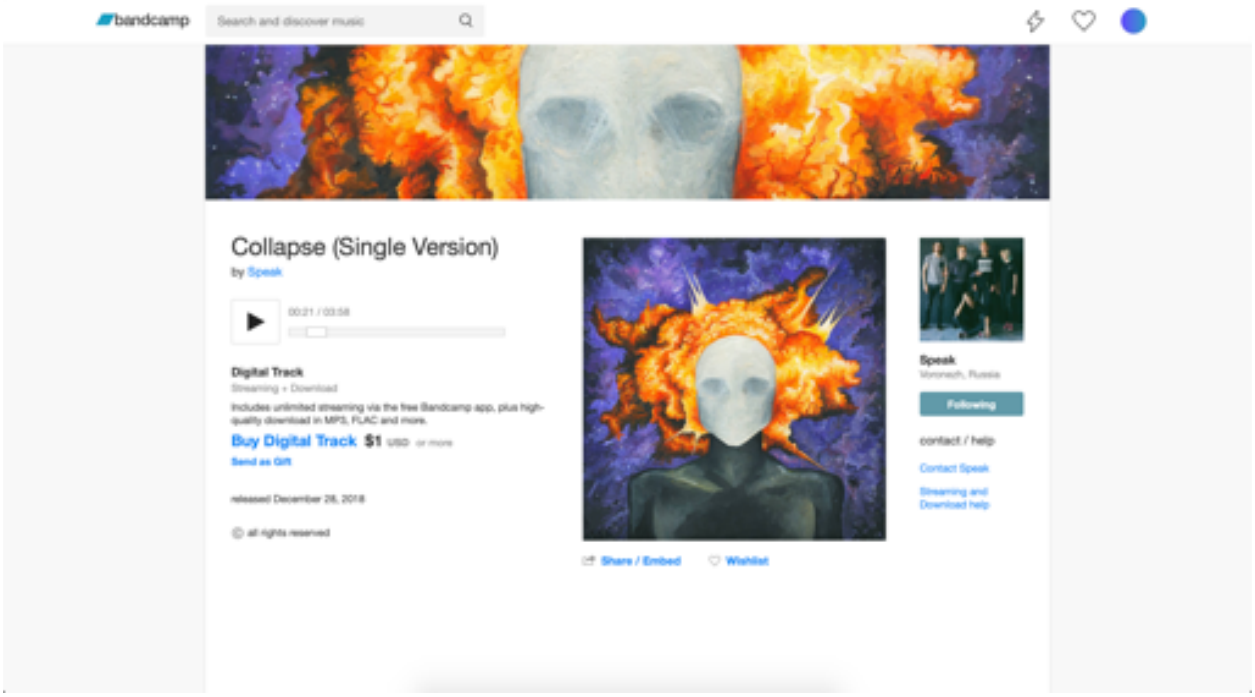

Рис. 1.8 Пример страницы музыкальной композиции в «Bandcamp»

1.3.2 Музыкальная платформа «Soundcloud»

Данный сервис представляет из себя более классический в представлении пользователей музыкальный проигрыватель. Также, как и в предыдущем рассмотренном аналоге, оценка и критика композиций мало отличается от стандартной модели, используемой не только в музыкальных приложениях, но и в социальных сетях, форумах, и т.д. Иллюстрация страницы музыкальной композиции представлена на рисунке 1.9.

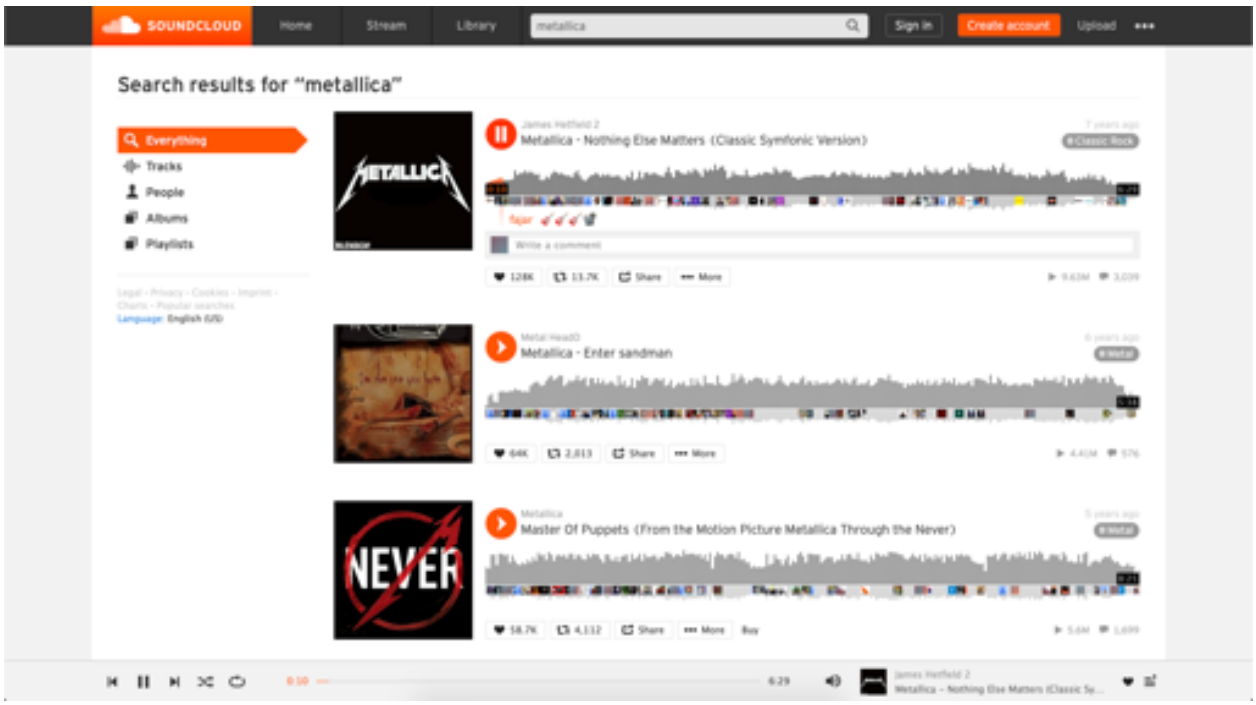

Рис. 1.9 Пример страницы музыкальной композиции в «Soundcloud»

1.3.3 Интернет-сервис потокового аудио «Spotify»

Данный сервис является главным музыкальным стриминговым сервисом в мире, предлагающим легальное бесплатное прослушивание музыки.

Его музыкальная база насчитывает более 50 млн. композиций, 248 млн. активных пользователей и 3 миллиарда пользовательских плейлистов. Иллюстрация страницы сервиса представлена на рисунке 1.10.

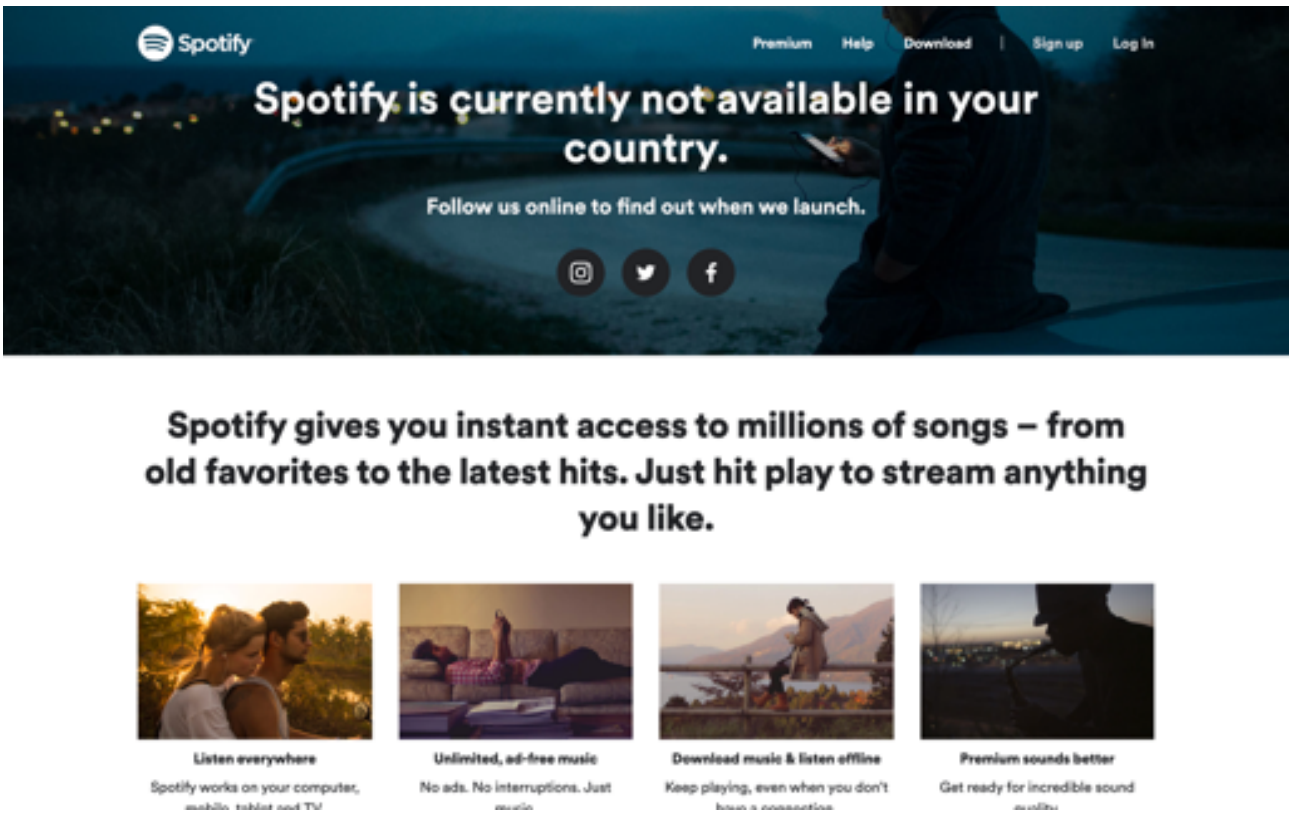

Рис. 1.10 Пример сервиса «Spotify»

На данный момент сервис предоставляет услуги преимущественно прослушивания музыки. Одна из его программ Spotify For Artists, предназначенная для выкладывания и продвижения авторской музыки, в текущее время находится в стадии беты.

1.3.4 Сервис распознавания потокового аудио «Shazam»

Shazam — бесплатный кроссплатформенный проект, позволяющий пользователю определить, что за песня играет в данный момент (ранее был доступен только коммерческий проект и только для мобильных устройств, в данный момент существует приложение для iOS, Android, macOS, WatchOS). Компания была основана в 1999 году в Лондоне.

Пользователь Shazam использует микрофон устройства для записи фрагмента музыки, которая играет где-либо. Затем программа сравнивает фрагмент с центральной базой данных и при успешном сопоставлении

выдаёт информацию о треке. В настоящий момент сервис предоставляет информацию о более чем 11 млн треков.

Shazam может идентифицировать записанные звуки, которые передаются из любых источников, при условии, что уровень фонового шума не слишком высок. Shazam хранит каталог аудио, опознанных при помощи программы, давая прямые ссылки на данные треки на YouTube и Apple Music, если таковые там могут быть найдены. Иллюстрация страницы сервиса представлена на рисунке 1.11.

Рис. 1.11 Пример сервиса «Shazam»

1.3.5 Аудиоплеер «Apple Music»

Apple Music — музыкальная служба, представленная компанией Apple на WWDC 2015. Предоставляет доступ к миллионам композиций из одной из самых крупных аудиобиблиотек в мире — iTunes Store.

В Apple Music есть функция радио, при включении которой подбираются музыкальные композиции в соответствии с предпочтениями слушателя.

Среди прочих особенностей сервиса присутствует возможность сохранения музыки на устройство для прослушивания офлайн. Иллюстрация страницы сервиса представлена на рисунке 1.12.

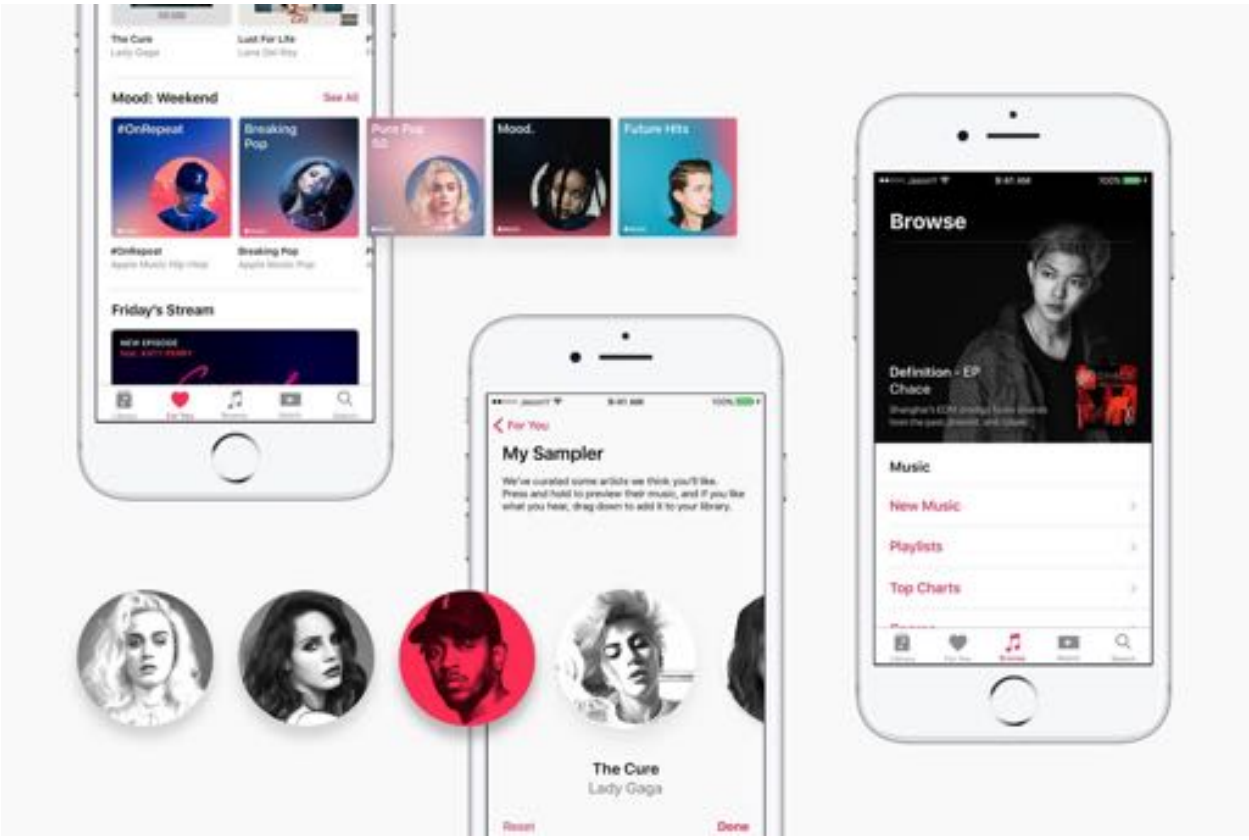

Рис. 1.12 Пример сервиса «Apple Music»

1.3.6 Результат исследования существующих аналогов

Данные сервисы созданы для выполнения определенных музыкальных задач и обладают различными достоинствами и недостатками. Одними из недостатков были выявлены:

− отсутствие развернутой конструктивной оценки качества музыкальных композиций;

− недостаточная точность или ограниченность возможностей работы моделей машинного обучения.

Работа над данными проблемами является одной из целей разрабатываемого проекта.

Список использованных источников

1. Методические указания по оформлению курсовых и выпускных квалификационных работ (факультет ПММ) / Т. В. Азарнова [и др.]. – Воронеж : Издательский дом ВГУ, 2019. – 48 стр.;

2. Музыка – URL: http://ru.wikipedia.org/wiki/Музыка (дата обращения: 20.01.2021).;

3. Анализ аудиоданных с помощью глубокого обучения и Python – URL: https://medium.com/nuances-of-programming/анализ-аудиоданных-с-помощью-глубокого-обучения-и-python-часть-1-2056fef8525e (дата обращения: 20.01.2021);

4. Акустическая фонетика – URL: https://myfilology.ru/russkiiyazyk/fonetika-i-fonologiia/akusticheskaia-fonetika-akusticheskie-priznaki-zvukov-russkogo-iazyka/ (дата обращения: 20.01.2021).