|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |
| **ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**  итоговой аттестационной работы    на тему:**Применение глубоких сверточных нейронных сетей в задаче классификации аудиоданных по гендерному признаку.**  Слушатель **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_Пермяков Н.П.\_\_\_\_\_\_\_\_**  (Подпись, дата) (И.О.Фамилия)  Руководитель **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_Гапанюк Ю.Е.\_\_\_\_\_\_\_\_**  (Подпись, дата) (И.О.Фамилия)  Нормоконтролер **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_Афанасьева Н.В.\_\_\_\_\_\_**  (Подпись, дата) (И.О.Фамилия)  2023 | |

**Задание на выпускную квалификационную работу**

|  |
| --- |
| **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)**  УТВЕРЖДАЮ  Директор ИСОТ  \_\_\_\_\_\_В.Г. Брекалов  « \_\_\_ » \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 20 \_\_ г.  **ЗАДАНИЕ**  **на выполнение итоговой аттестационной работы**  Слушатель \_\_\_\_\_\_\_\_\_\_\_\_Пермяков Никита Петрович\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  (фамилия, имя, отчество)  Тема итоговой аттестационной работы **Применение глубоких сверточных нейронных сетей в задаче классификации аудиоданных по гендерному признаку**.  Тема итоговой аттестационной работы утверждена распоряжением по ИСОТ МГТУ им. Н.Э. Баумана № \_\_\_\_\_\_\_\_\_\_\_от « \_\_\_ » \_\_\_\_\_\_\_\_\_\_\_\_\_ 20\_\_ г.  Краткое содержание итоговой аттестационной работы:  Выпускная квалификационная работа на тему «Применение глубоких сверточных нейронных сетей в задаче классификации аудиоданных по гендерному признаку» включает в себя обучение модели для определения голоса, с последующим использованием в разработанном приложении. |

|  |
| --- |
| ***Оформление итоговой аттестационной работы:***  Расчетно-пояснительная записка на 48 листах формата А4.  Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т.п.)  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Дата выдачи задания « \_\_\_ » \_\_\_\_\_\_\_\_\_\_\_\_ 20\_\_ г.  В соответствии с учебным/календарным планом итоговую аттестационную работу выполнить в полном объеме в срок до « \_\_\_\_ » \_\_\_\_\_\_\_\_\_\_\_\_ 20\_\_\_\_ г.  **Руководитель**  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ /\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  (Подпись, дата) (И.О.Фамилия)  **Слушатель \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  (Подпись, дата) (И.О.Фамилия) |

# АННОТАЦИЯ

Расчётно-пояснительная записка квалификационной работы бакалавра содержит 48 страниц. Работа включает в себя 60 иллюстраций. В процессе выполнения было использовано 30 источников.

Объектом разработки является пользовательское веб - приложение на фреймворке Flask.

Цель работы заключается разработка модели, способной с высокой долей точности проводить классификацию аудиоданных по гендерному признаку.

В процессе выполнения квалификационной работы бакалавра была успешно применена глубокая сверточная нейронная сеть и решена задача бинарной классификации аудиоданных. Разработанная модель определяет гендер по голосу с точностью 95%.

# СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 6](#_Toc147052088)

[1 АНАЛИТИЧЕСКАЯ ЧАСТЬ 8](#_Toc147052089)

[1.1 Постановка задачи 8](#_Toc147052090)

[1.2 Описание используемых методов. 8](#_Toc147052091)

[1.3 Разведочный анализ данных 15](#_Toc147052092)

[2 ПРАКТИЧЕСКАЯ ЧАСТЬ 16](#_Toc147052093)

[2.1 Предобработка данных 16](#_Toc147052094)

[2.2 Разработка и обучение модели 31](#_Toc147052095)

[2.3 Тестирование модели 36](#_Toc147052096)

[2.4 Проверка работы модели на собственном голосе 39](#_Toc147052097)

[2.5 Разработка приложения 40](#_Toc147052098)

[ЗАКЛЮЧЕНИЕ 45](#_Toc147052100)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 45](#_Toc147052101)

# ВВЕДЕНИЕ

Звук — это физический процесс, который заключается в распространении упругих механических волн в различных средах, таких как твёрдые, жидкие или газообразные.

Речь человека состоит из совокупности отдельных звуков, которые создаются голосовыми связками. Звуки, не видимые глазом, воспринимаются слухом человека благодаря воздействию на его барабанные перепонки.

Термин "аудиосигнал" используется для описания колебаний, соответствующих речи, музыке и другим звуковым проявлениям в слышимом диапазоне частот.

На сегодняшний день идут активные исследования в области аудиосигналов. В повседневной жизни мы видим голосовых ассистентов, способных распознавать и обрабатывать речь.

Аудиосигнал может быть представлен в виде массива данных, который легко визуализировать в виде графика колебаний, пересекающего нулевую линию.

В данном исследовании мы рассмотрим применение глубоких сверточных нейронных сетей для классификации аудиоданных по гендерным признакам.

Эта тема представляет интерес для банковской сферы и компаний, которые управляют контактными центрами. В настоящее время клиенты могут проводить финансовые операции по телефону без необходимости предъявления физического документа, удостоверяющего личность. Однако возникли случаи мошенничества, когда злоумышленники, зная личные данные клиента и кодовые слова, представляются им при звонке в контактный центр банка. Если бы существовала проверка голосового анализа, операторы контактных центров могли бы установить разницу между голосами клиента и мошенника. Это помогло бы снизить число случаев мошенничества.

Целью этого исследования является разработка модели, способной высоко точно классифицировать аудиоданные по гендерным признакам. Для демонстрации работы модели мы создадим простое в использовании веб-приложение на основе фреймворка Flask.



Рисунок 1 – Скриншот дизайна работающего приложения на фреймворке Flask

1. **АНАЛИТИЧЕСКАЯ ЧАСТЬ**

**1.1 Постановка задачи**

Для нашего исследования мы воспользуемся набором данных TIMIT. Этот набор включает в себя аудиозаписи носителей восьми основных диалектов американского английского языка. Для каждого диктора доступно по десять фраз, фонетически разнообразных. Некоторые из фраз одинаковы для всех дикторов, в то время как другие отличаются. В TIMIT содержится 438 аудиозаписей мужчин и 192 аудиозаписи женщин. В обучающем наборе данных насчитывается 4620 аудиозаписей, а в тестовом — 1680 аудиофайлов.

Мы начнем с демонстрации процесса предобработки аудиоданных, используя один из аудиофайлов в качестве примера. Затем проведем анализ набора данных TIMIT.

Для обучения классификатора пола по голосу мы создадим класс timit\_dataloader, который позволит извлекать метки пола, возраста и идентификаторы дикторов, а также предоставит методы предобработки аудиосигнала.

Далее мы разработаем архитектуру нейронной сети и проведем обучение модели. Мы также используем метрики для оценки качества классификации и выберем модель, которая покажет наилучший результат.

В конечном итоге, для продемонстрации работы нашей модели, мы создадим приложение с графическим интерфейсом. С помощью этого приложения пользователи смогут записать голос и проверить, насколько точно определяется пол по аудиосигналу.

**1.2 Описание используемых методов**

Речевой сигнал должен быть преобразован в определенный тип параметрического представления для дальнейшего анализа и обработки. Существует достаточно большое количество методов для представления речевого сигнала в виде вектора признаков. Например, методы, такие как Linear Prediction Coding (LPC) и Mel-Frequency Cepstral Coefficients (MFCC), предоставляют разнообразные подходы к этой задаче. Среди них MFCC, или метод мел-частотных кепстральных коэффициентов, является наиболее известным и широко используемым методом, и именно его мы будем использовать в данной работе.

Этот метод основан на изменении человеческого голоса с использованием частотных треугольных фильтров с критической пропускной способностью. Эти фильтры размещены на разных частотах, чтобы получить важные характеристики для анализа речи. Размеры окон выбираются на основе мел-шкалы: первое окно очень узкое и охватывает низкие частоты, где звуки менее различимы, и размер окна увеличивается с увеличением частоты для адаптации к человеческому восприятию. MFCC обычно рассчитываются с использованием треугольных фильтров, снабженных центральными частотами, линейно расположенными для низких частот и логарифмически для высоких частот. Параметры каждого фильтра зависят от центральных частот двух соседних фильтров и их количества в мел-шкале.

Человеческое восприятие звуков в речи не линейно по частоте, и поэтому для представления сигнала в мел-шкале мы используем формулу (1), которая описывает эту зависимость:

, (1)

где 𝑓 – частоты по обычной (линейной) шкале;

𝐻(𝑓) – частоты по мел-шкале.

Энергия, попадающая в каждое из окон анализа, рассчитывается путем умножения энергетического спектра сигнала на оконную функцию, как показано в формуле (2):

, (2)

где 𝑥𝑚 – энергетический коэффициент;

𝑚 = 1, … , 𝑀 – количество фильтров;

𝑋𝑘 – амплитудные коэффициенты спектра сегмента;

𝐻𝑚(𝑓) – функция.

Результатом этого вычисления является набор коэффициентов , содержащих спектральную информацию речевых сегментов. После вычисления энергии, мы применяем логарифмирование коэффициентов, так как человеческое восприятие громкости воспринимает нелинейно: удвоение воспринимаемой громкости требует увеличения энергии в 8 раз. Для этого мы используем формулу (3):

, (3)

Поскольку коэффициенты спектра mel являются действительными числами, мы можем преобразовать их во временной интервал, используя дискретное косинусное преобразование 4:

, (4)

Шкала Мел - это способ оценки восприятия частоты звука человеческим ухом, который учитывает, что люди лучше различают небольшие изменения частот на низких частотах, чем на высоких. Эта зависимость не является линейной и может быть описана формулой 5:

, (5)

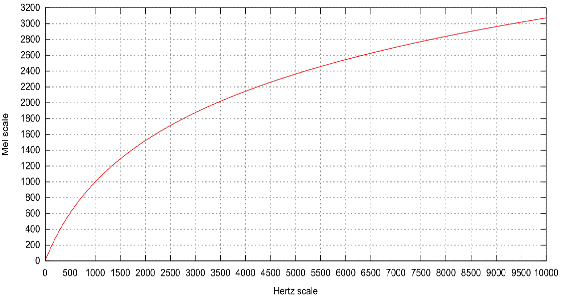


Рисунок 2 – Шкала Мел

Вычисление мел-частотных кепстральных коэффициентов включает в себя следующие этапы:

Разбиение исходного аудиосигнала на фреймы. Обычно размер фрейма выбирается в диапазоне от 20 до 40 миллисекунд, так как в этом временном интервале предполагается, что речевой сигнал изменяется незначительно.

Речевой сигнал не является периодическим и имеет конечную длительность, что может привести к эффекту "утечки" при применении преобразования Фурье. Для уменьшения влияния этого эффекта на результат, каждый фрейм умножается на оконную функцию Хэмминга.

Вычисление периодограммы для каждого фрейма, что представляет собой спектральную мощность сигнала.

Применение мел-фильтров. Для этого треугольные фильтры (обычно от 20 до 40) умножаются на периодограмму и суммируются. Это позволяет получить энергии в разных мел-полосах.

Полученные энергии логарифмируются. Это согласуется с восприятием громкости человеческим слухом, так как человеческое восприятие громкости не линейно. Чтобы удвоить воспринимаемую громкость звука, необходимо увеличить энергию в 8 раз. Это приводит к сжатию динамического диапазона.

Затем, с использованием дискретного косинусного преобразования, получаем мел-частотные кепстральные коэффициенты, которые представляют собой важные характеристики аудиосигнала и используются для дальнейшего анализа и обработки.

Визуально процесс представлен ниже на рисунках.

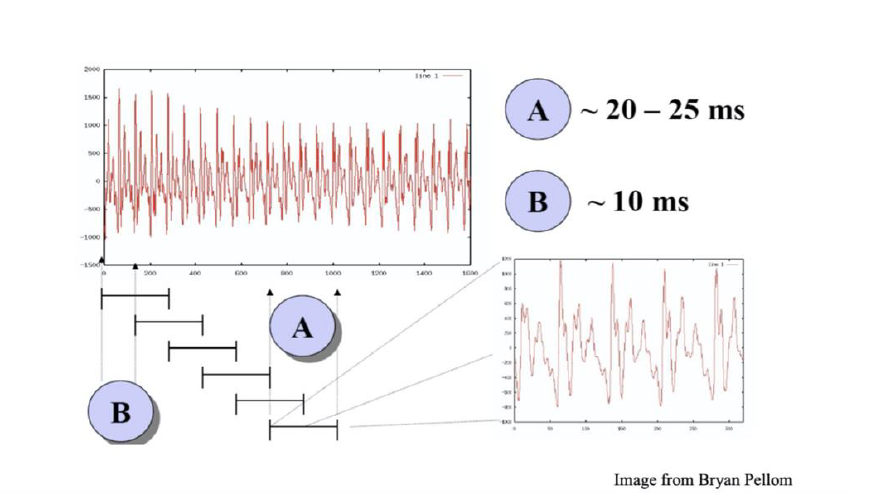


Рисунок 3 – Разбиение на фреймы

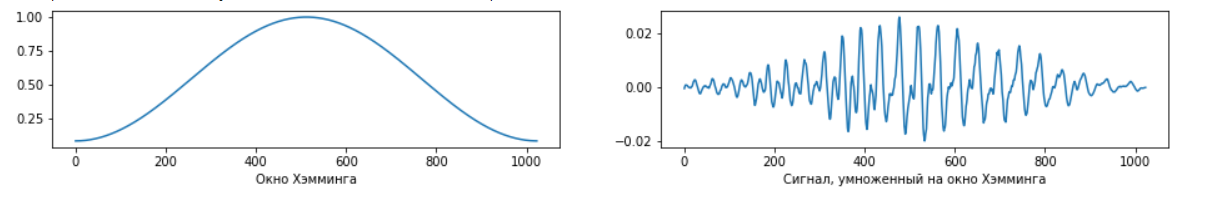


Рисунок 4 – Сигнал, умноженный на окно Хэмминга

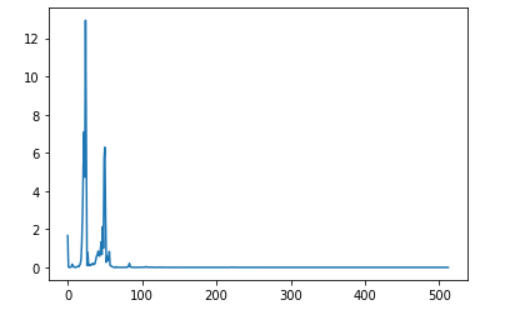


Рисунок 5 – Периодограмма

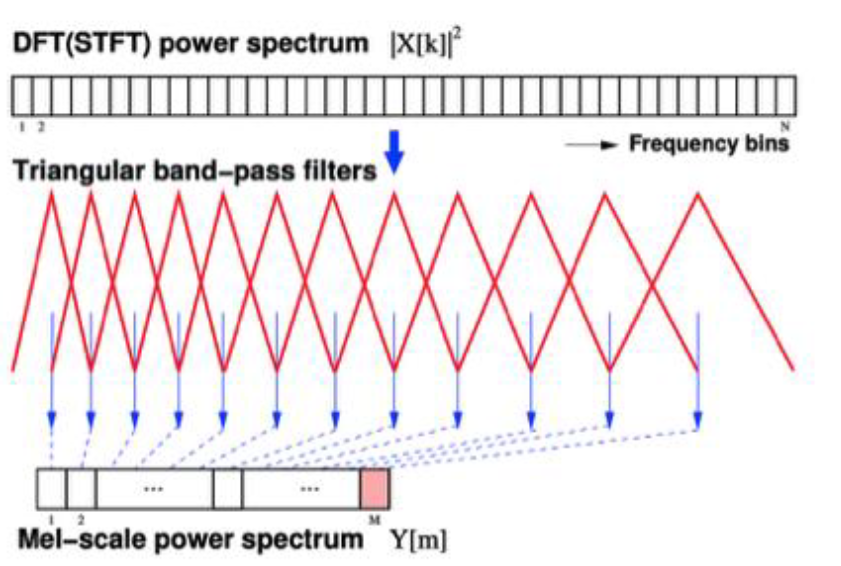


Рисунок 6 – Блок мел-фильтров

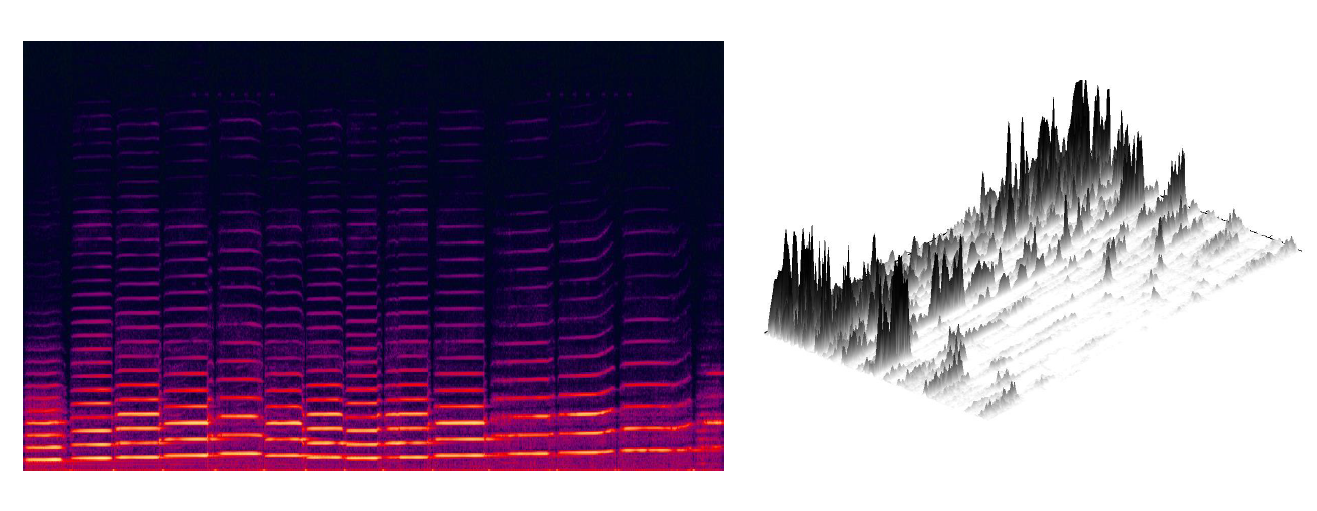


Рисунок 7 – Мел- Кепстральные коэффициенты

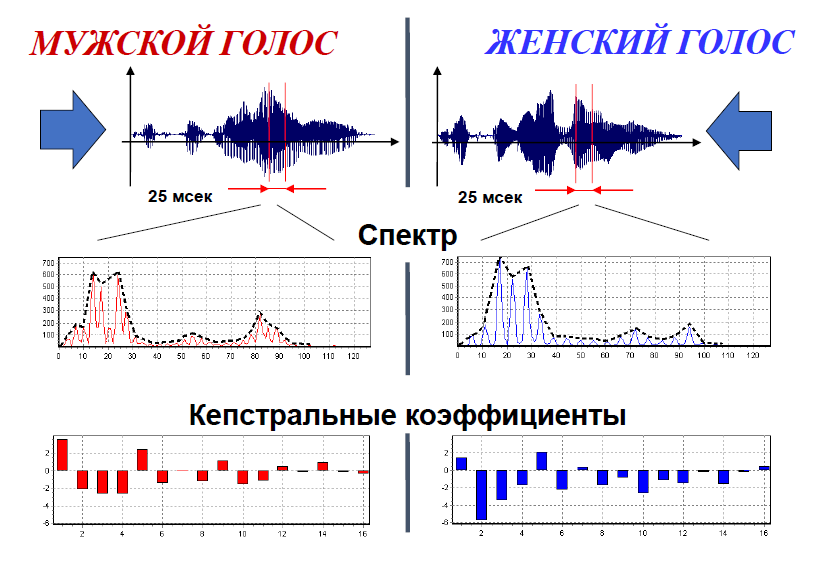


Рисунок 8 – Сравнение мужского и женского голоса

* 1. **Разведочный анализ данных**

Прежде чем передавать данные в работу моделей машинного обучения, важно провести обработку и очистку данных. Очевидно, что "грязные" и необработанные данные могут содержать искажения и пропущенные значения, что является ненадежным и может привести к неверным результатам при моделировании. Однако, без обоснования, удаление данных также не является правильным подходом. Именно поэтому первым шагом необходимо изучить набор данных.

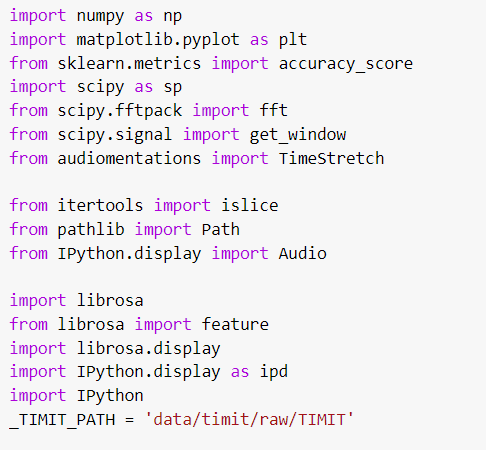
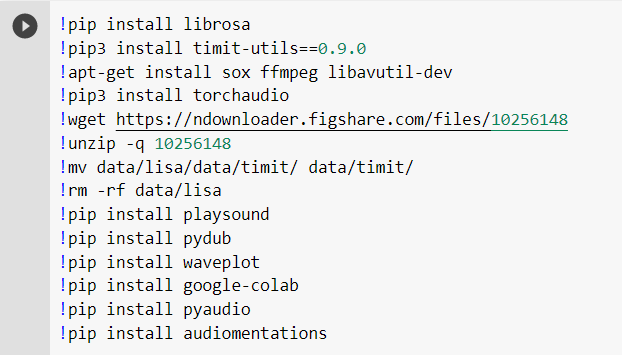
Цель разведочного анализа заключается в получении первоначальных представлений о характерах распределений переменных в исходном наборе данных, оценке качества данных (выявление наличия пропусков и выбросов), а также в выявлении взаимосвязей между переменными для последующего формулирования гипотез о наилучших моделях машинного обучения для решения задачи.

Для аудиоданных проведение разведочного анализа отличается от типовых наборов данных с табличной структурой. Например, набор аудиофайлов TIMIT характеризуется отсутствием шумов, стабильным темпом речи и записью профессиональными дикторами.

Для демонстрации основных методов предобработки аудиоданных мы осуществили аугментацию данных. Для искажения аудиозаписей было применено ускорение и замедление. После этого, для очистки аугментированных данных, мы провели мел-кепстральный анализ и применили методы очистки данных. На основе этой предобработки была построена модель нейронной сети.

Анализ проводился как для отдельных аудиозаписей, так и для всего набора данных TIMIT с целью лучшего понимания характеристик данных и выбора оптимальных методов обработки и моделирования.

1. **ПРАКТИЧЕСКАЯ ЧАСТЬ**
   1. **Предобработка данных**
2. Произведем установку библиотек, необходимых нам для исследовательской работы.



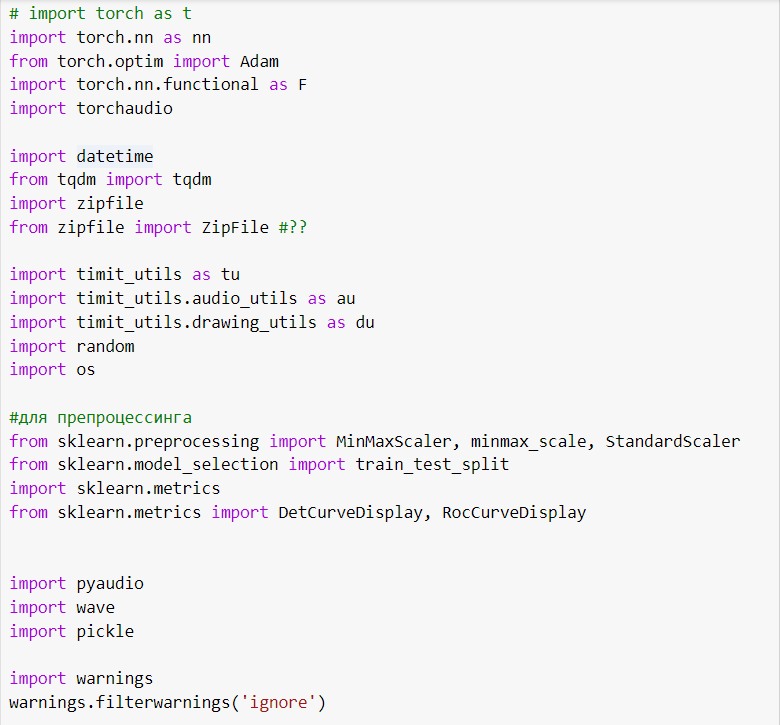
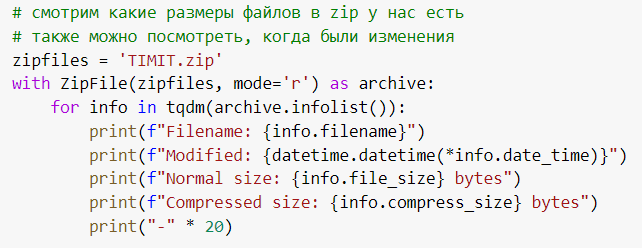
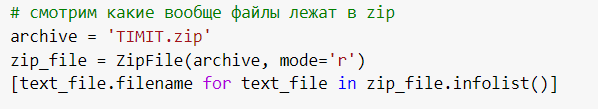
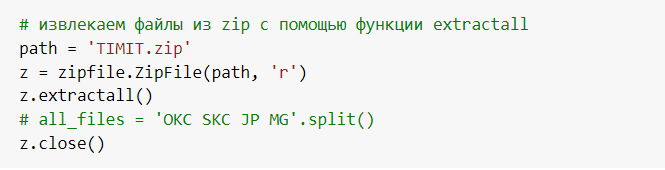


Рисунок 9 – Установленные библиотеки

1. Загрузим файлы и проверим какие файлы находятся в zip архиве.







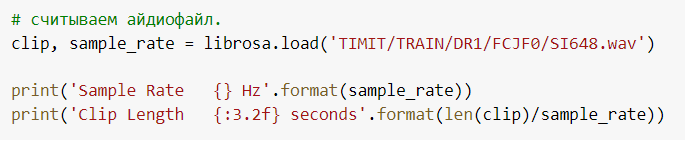
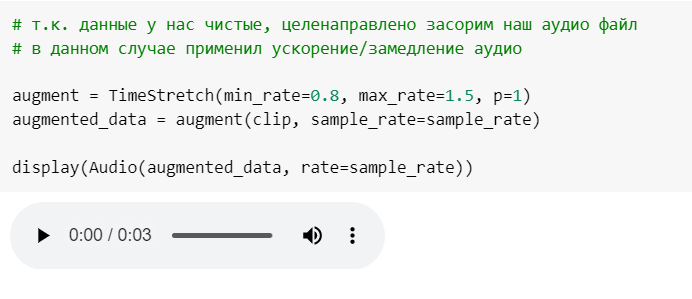
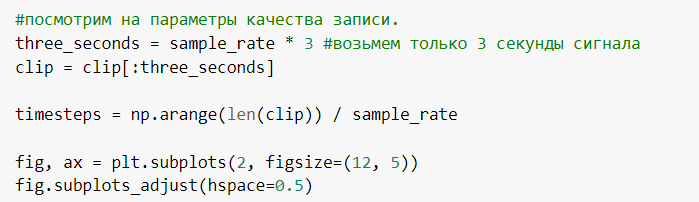
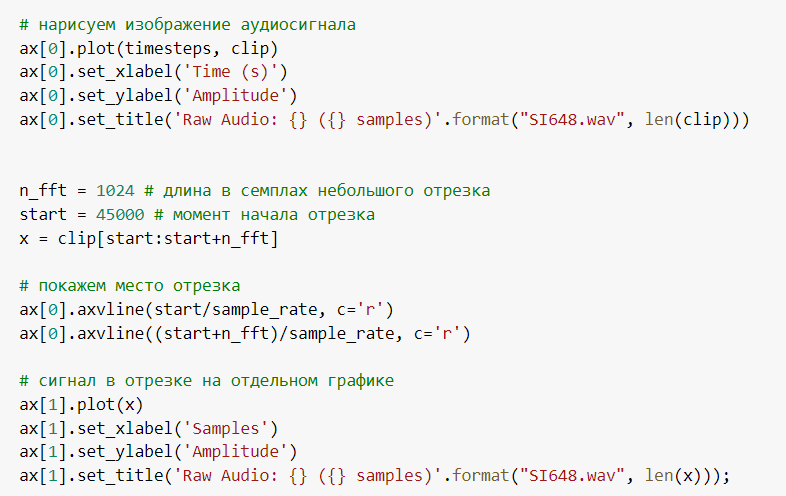


Рисунок 10 – Загрузка и проверка файлов

1. Предобработка данных на примере аудиофайла.







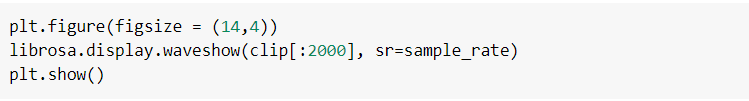
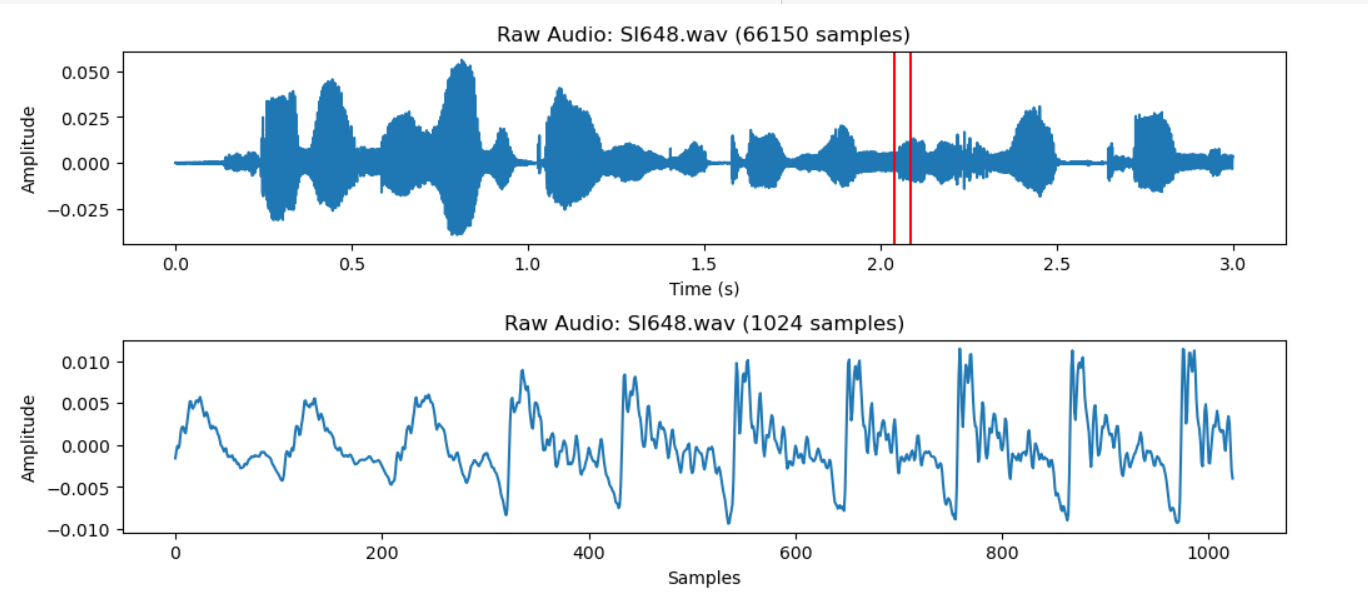


Рисунок 11 – Пример кода для предобработки

Результатом работы кода будут следующие графики:



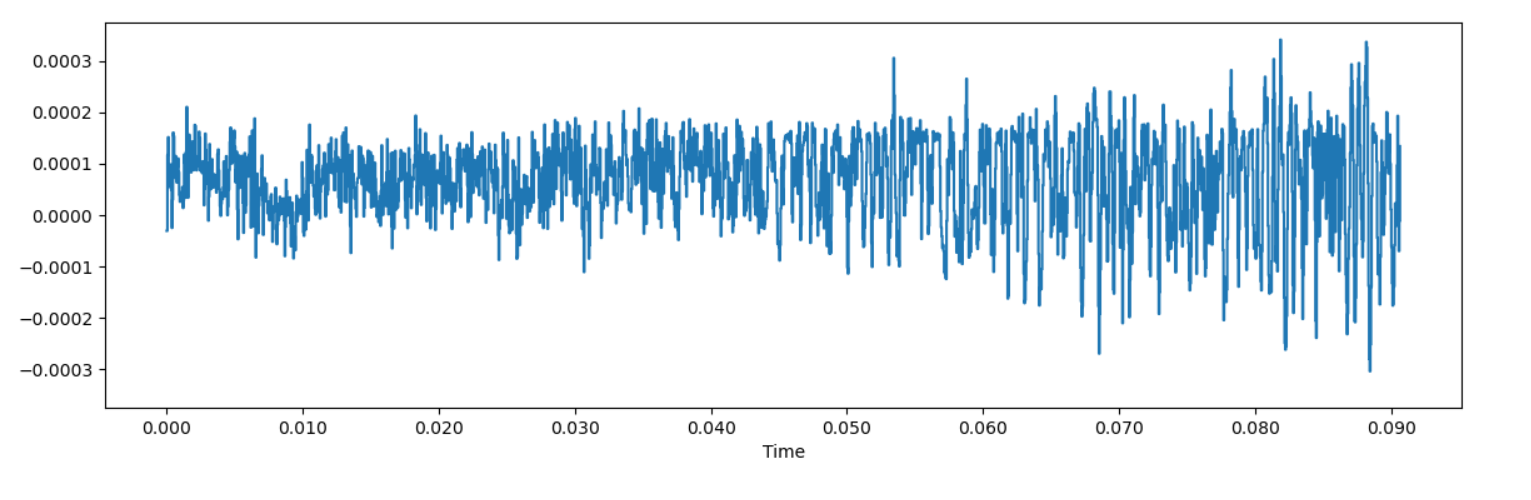


Рисунок 12 – Графики

Проведем параметризацию.

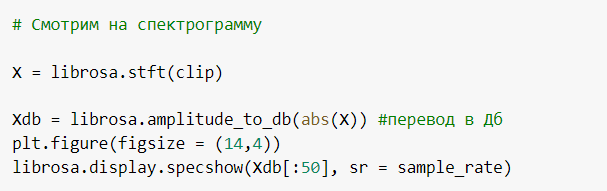


Рисунок 13 – Пример кода параметризации

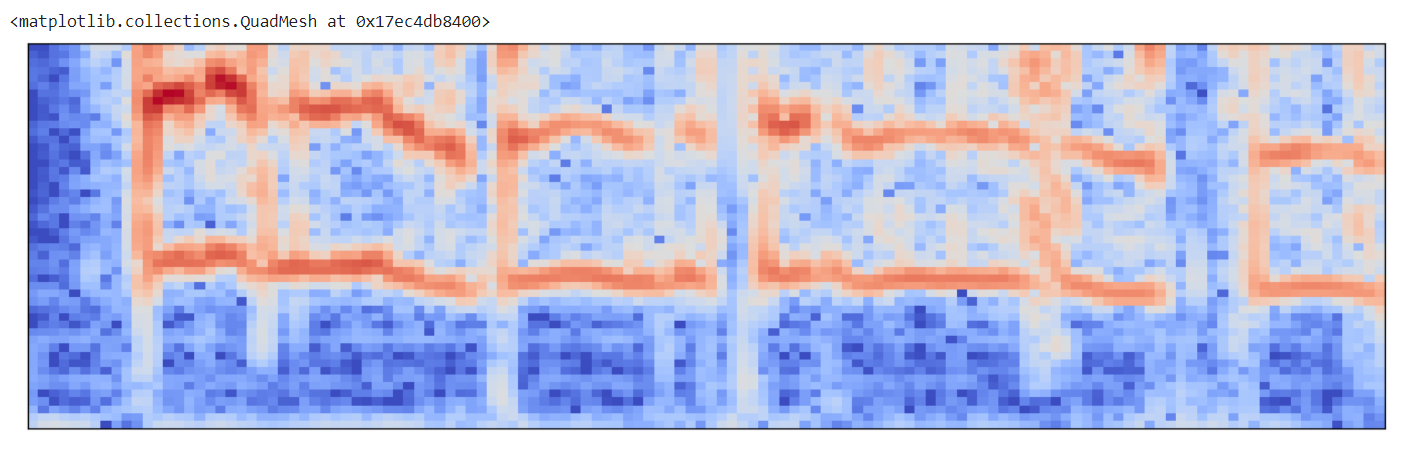


Рисунок 14 – Спектрограмма



Рисунок 15 – Пример кода

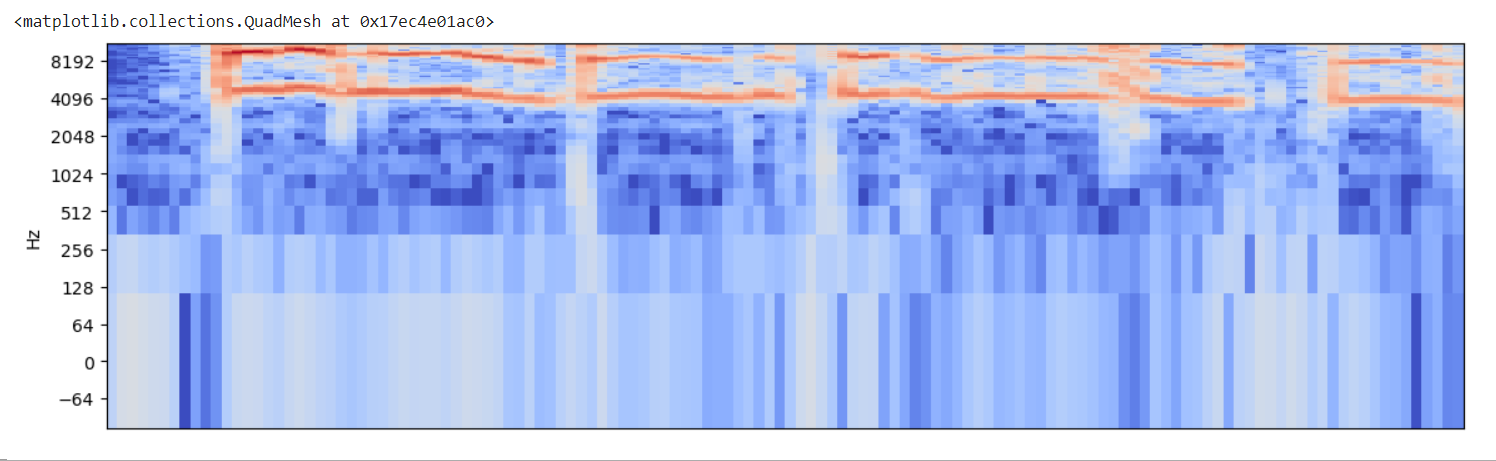


Рисунок 16 – Спектрограмма логарифмированная

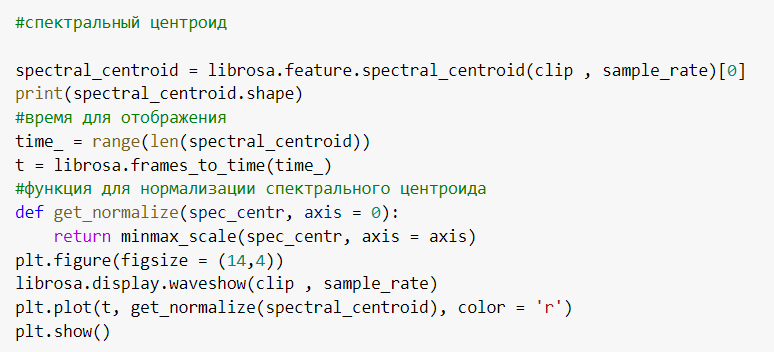


Рисунок 17 – Пример кода

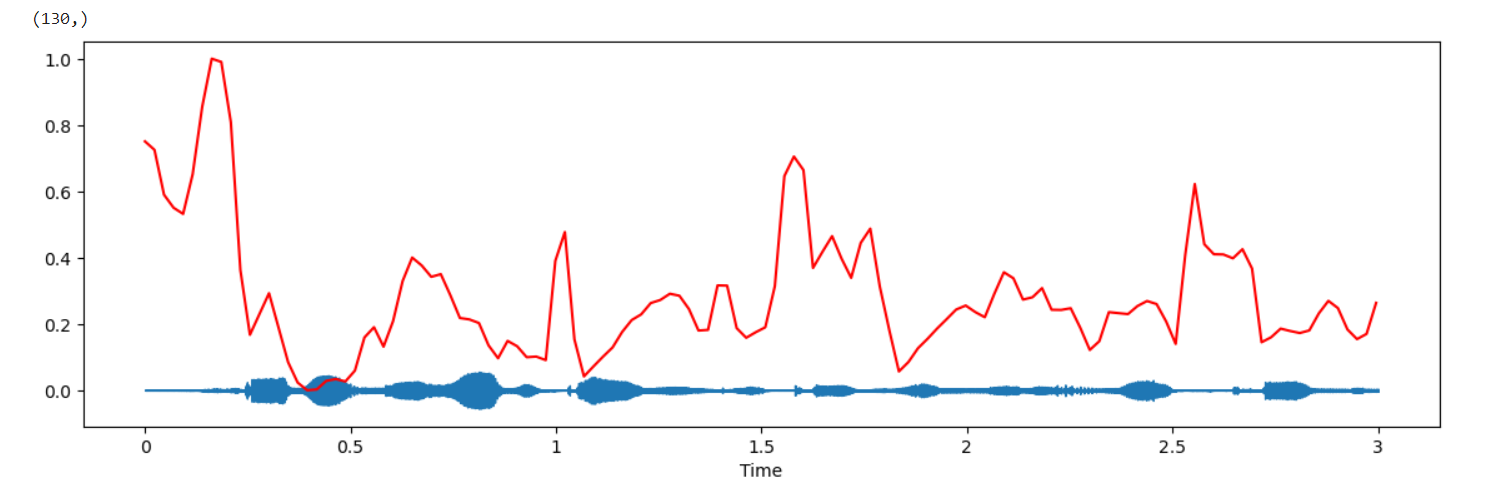


Рисунок 18 Спектральный центроид

Для получения признаков из аудиосигнала, которые затем можно подавать на классификатор, вычислим значения мел-кепстральных коэффициентов для данного сигнала.

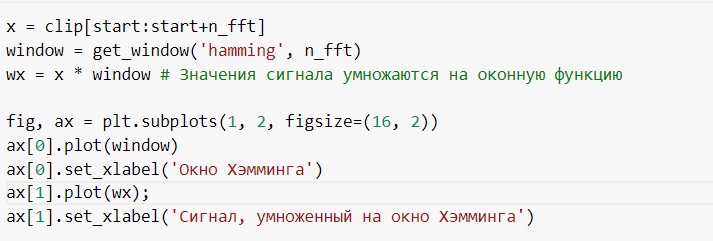


Рисунок 19 – Пример кода вычисления мел-кепстральных коэффициентов

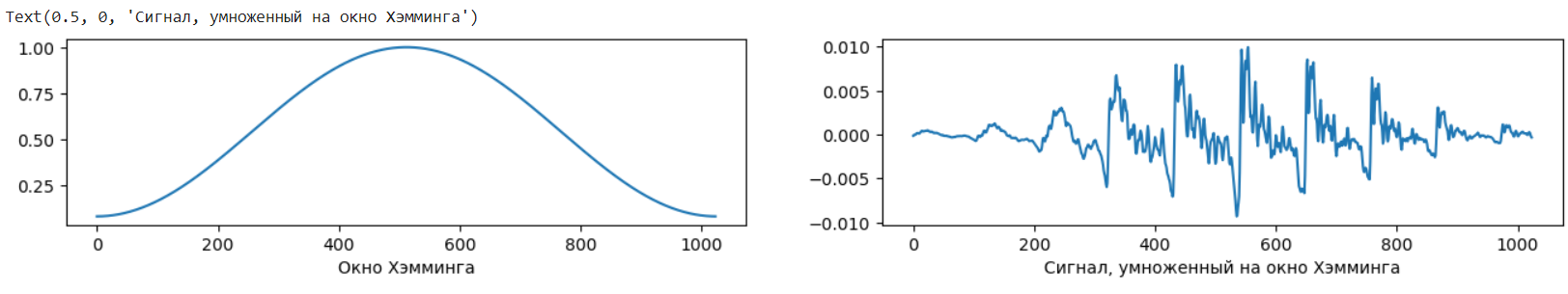


Рисунок 20 – Графики

Используем быстрое преобразование Фурье в выбранном окне.

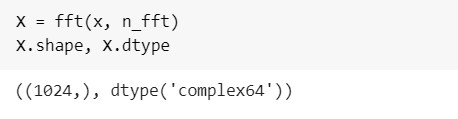


Рисунок 21 – Пример кода с результатом его работы.

Вычисляем периодограмму

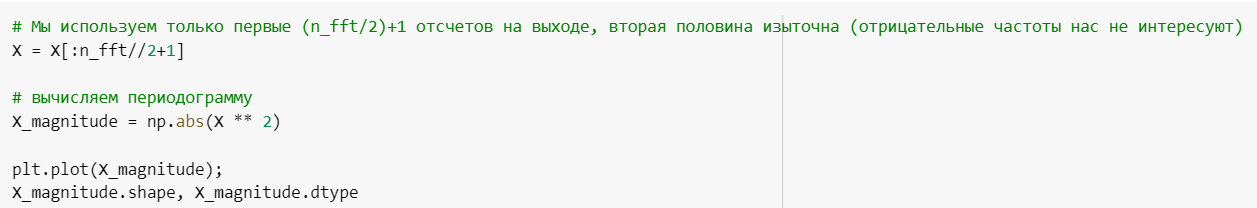


Рисунок 22 – Пример кода вычисления периодограммы

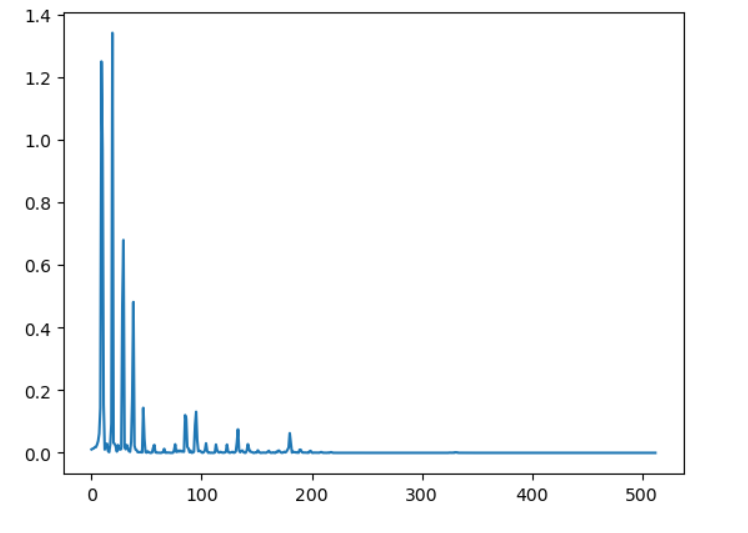


Рисунок 23 – График

Вычисляем блок мел-фильтров. Для этого треугольные фильтры (от 20 до 40) умножаются на периодограмму и суммируются. В результате мы получим энергии набора фильтров. Фильтры собираются в области низких частот, обеспечивая более высокое "разрешение" там, где это необходимо для распознавания.

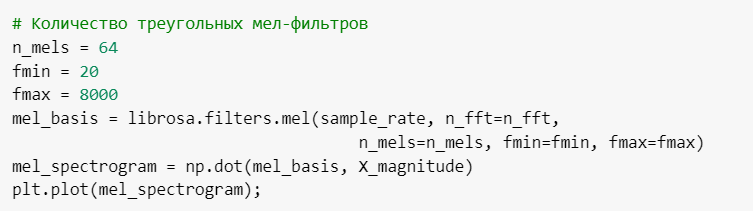


Рисунок 24 – Пример кода вычисления блок мел-фильтров

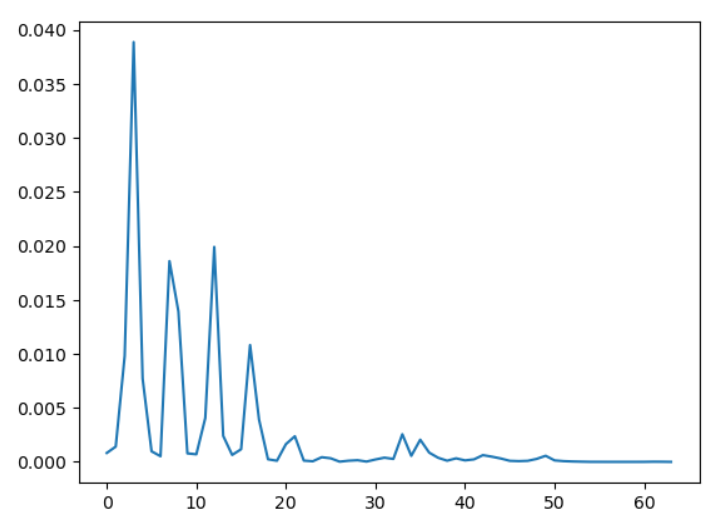


Рисунок 25 – График

Логарифмируем полученные значения.

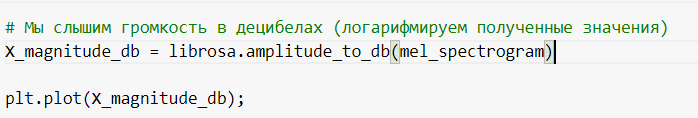


Рисунок 26 – Пример кода для логарифмирования

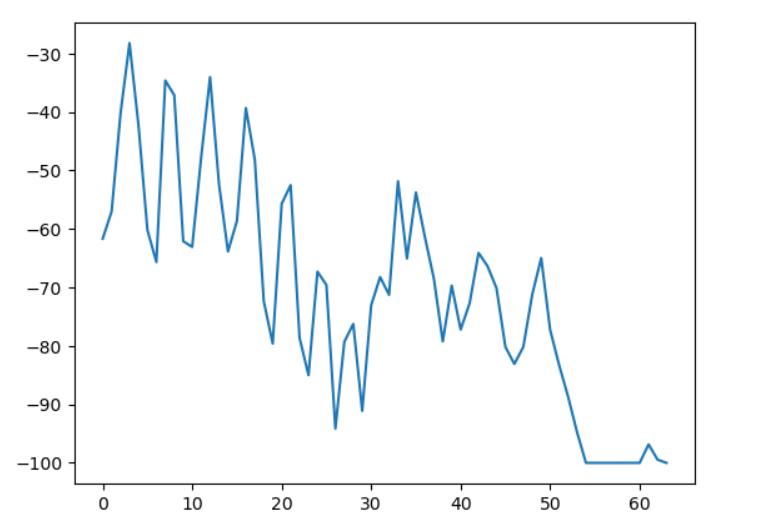


Рисунок 27 – График

Применить ко всему сигналу, а не к одному окну и получим следующую картину:



Рисунок 28 – Пример кода

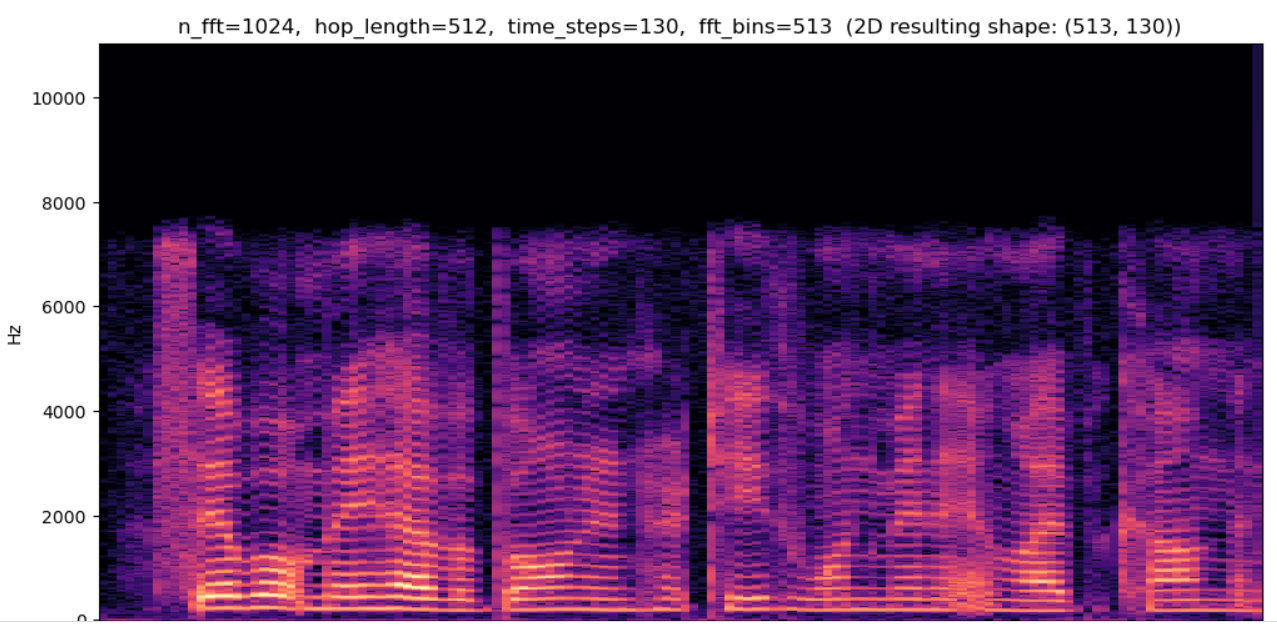


Рисунок 29 – График

Применим дискретное косинусное преобразование

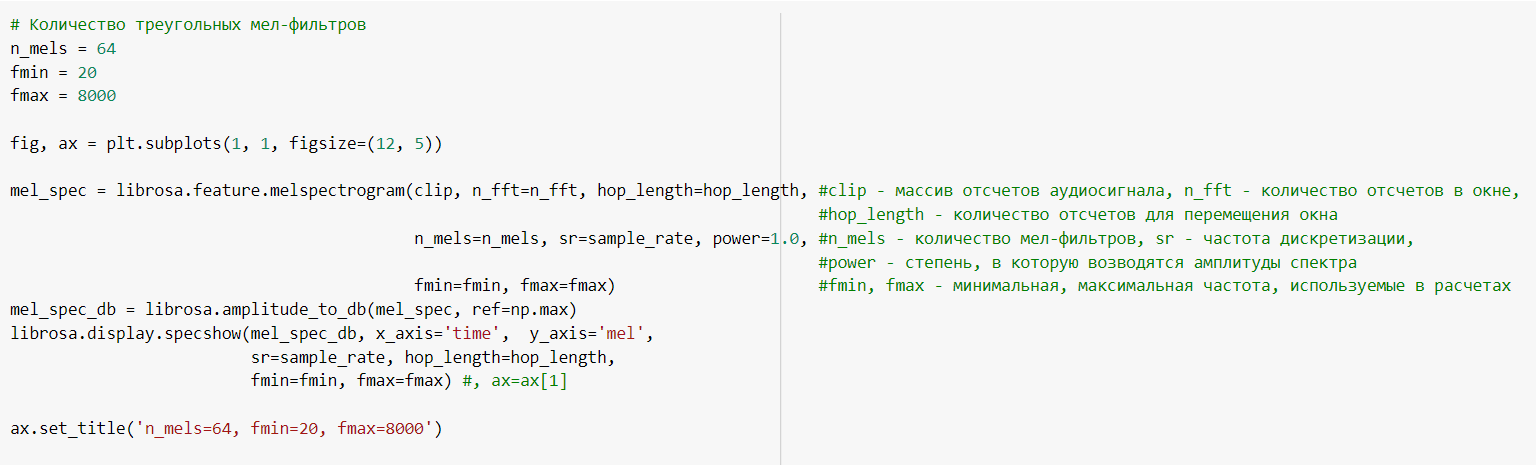


Рисунок 30 – Пример кода дискретного косинусного преобразования

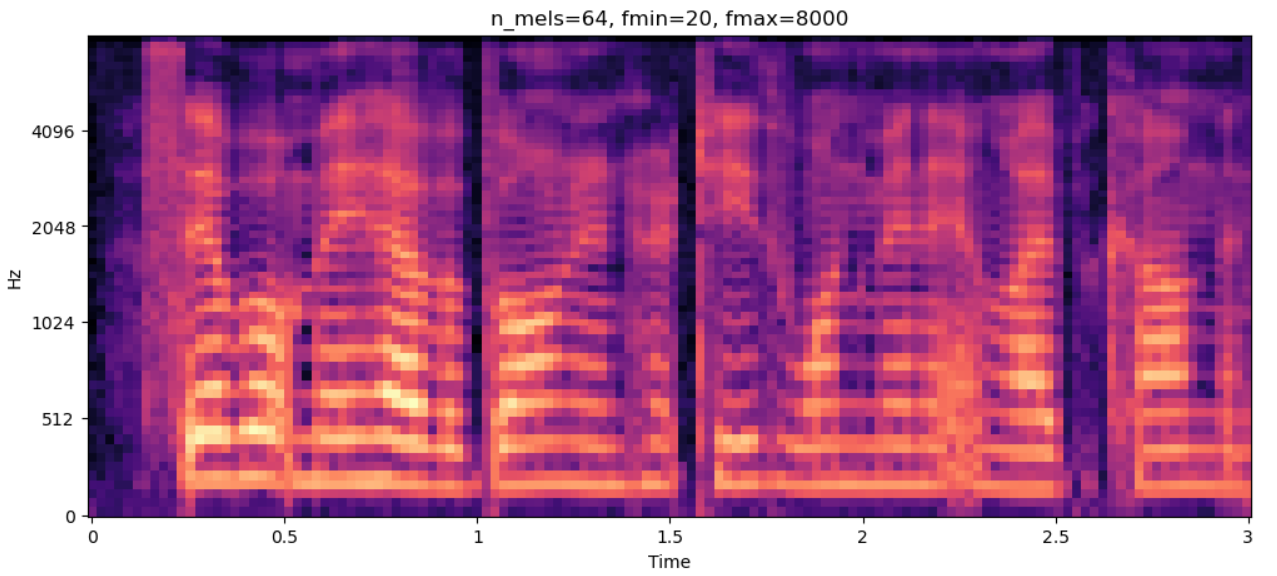


Рисунок 31 – График

Загрузим наши данные, содержащиеся в наборе TIMIT в корпус. Он содержит два подкорпуса, обучающую и тестовую выборку.

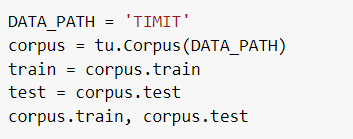


Рисунок 32 – Пример кода загрузки

Отобразим аудио предложения, слова и транскрипций фонем.



Рисунок 33 – Пример кода с результатом выполнения отображения

Создадим Датафрейм, содержащий количество слов, произнесенных дикторами. Речь в наборе данных на английском языке. Артикль «a» характерен для английского языка, по его наличию можно проверить были ли пропуски. Убедимся, что в нашем случае пропуски отсутствуют.

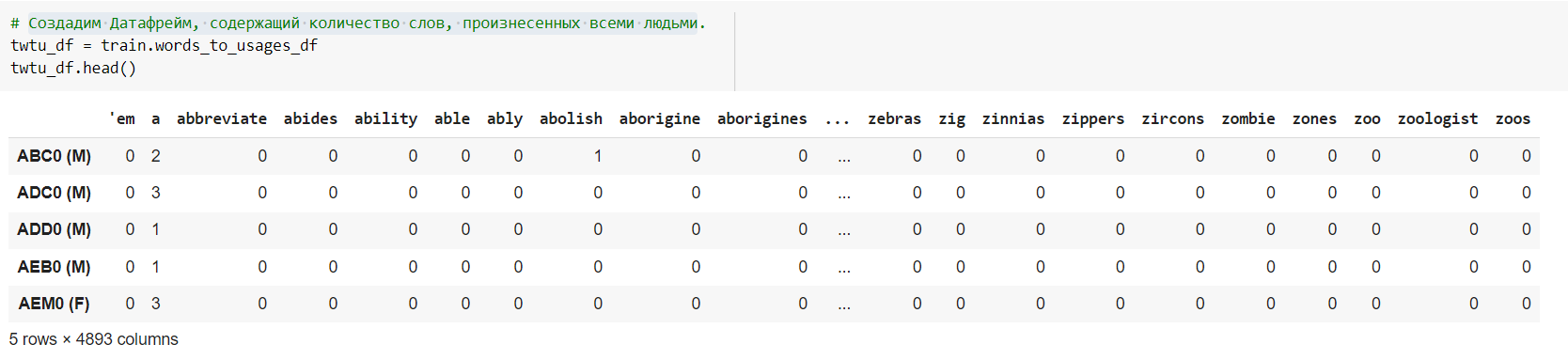
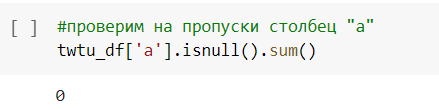


Рисунок 34 – Пример кода с результатом выполнения

Нас интересует столбец а из рисунка 34.



Пропуски отсутствуют. Создадим Датафрейм, содержащий количество фонем, произнесенных всеми людьми.

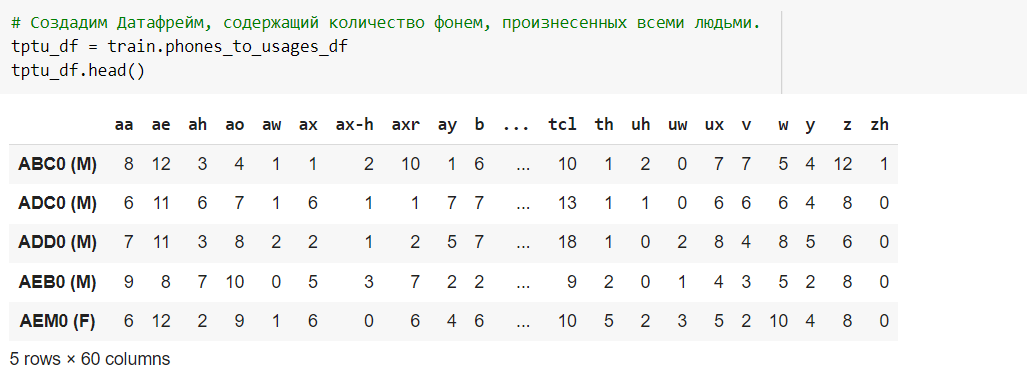


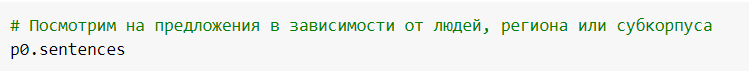
Рисунок 35 – Пример применённого кода с результатом выполнения.

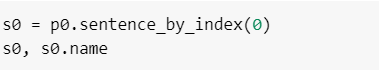


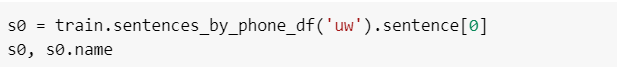
Рисунок 36 – Пример кода с результатом выполнения

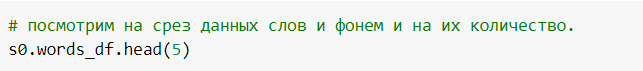
Изучим региональную составляющую по имеющемуся набору данных.



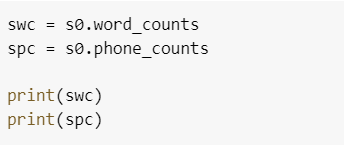












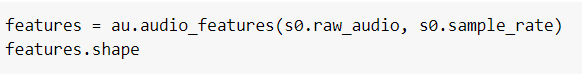
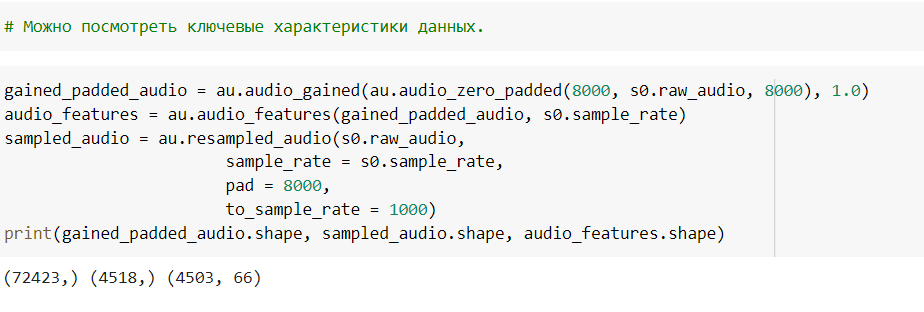


Рисунок 37 – Примеры кода для данной операции

Посмотрим ключевые характеристики данных.



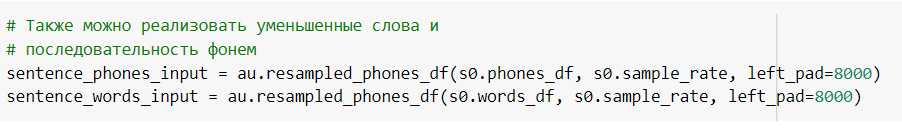


Рисунок 38 – Примеры кода с характеристиками

Визуализируем полученные данные с помощью графика.

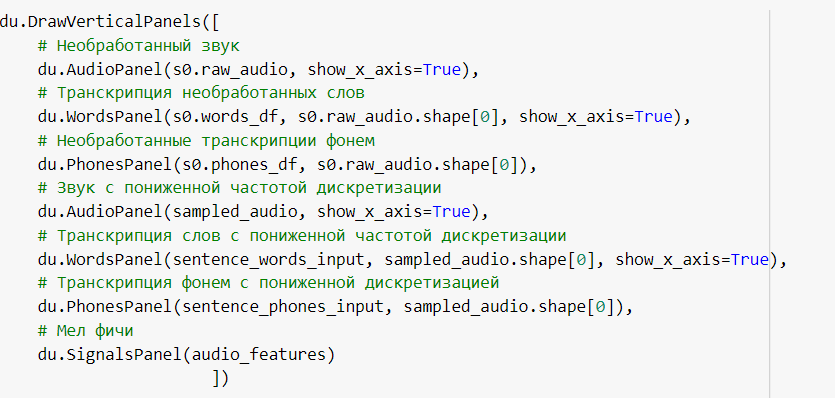


Рисунок 39 – Пример кода визуализации



Рисунок 40 – Полученное изображение

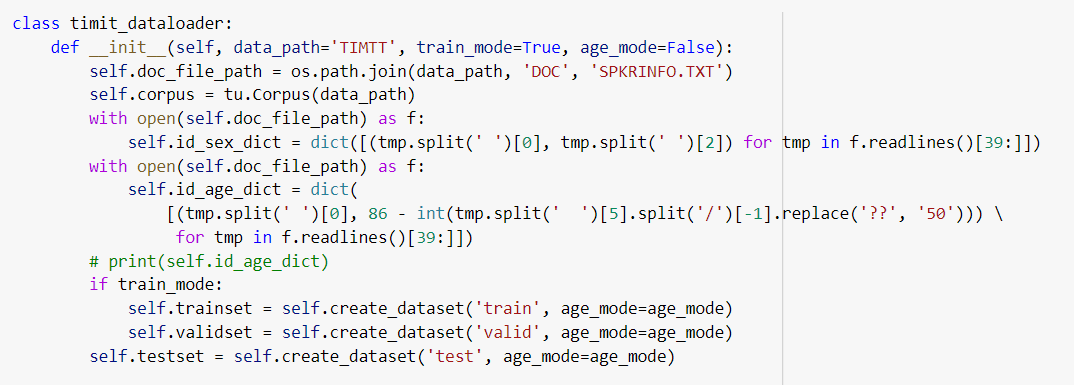
* 1. **Разработка и обучение модели**

Анализ, проведенный ранее, позволяет выявить частоты слов и фонем, характерные для мужчин и женщин. Эта информация будет полезной для разработки нейронной сети, которая будет способна определять пол говорящего по его голосу.

Для обучения классификатора пола по голосу мы создадим класс timit\_dataloader, который будет выполнять следующие функции: извлекать метки пола, возраста и идентификатора диктора из аудиоданных, а также предоставлять методы для предобработки аудиосигнала.



Рисунок 41 – Пример кода для текущей операции





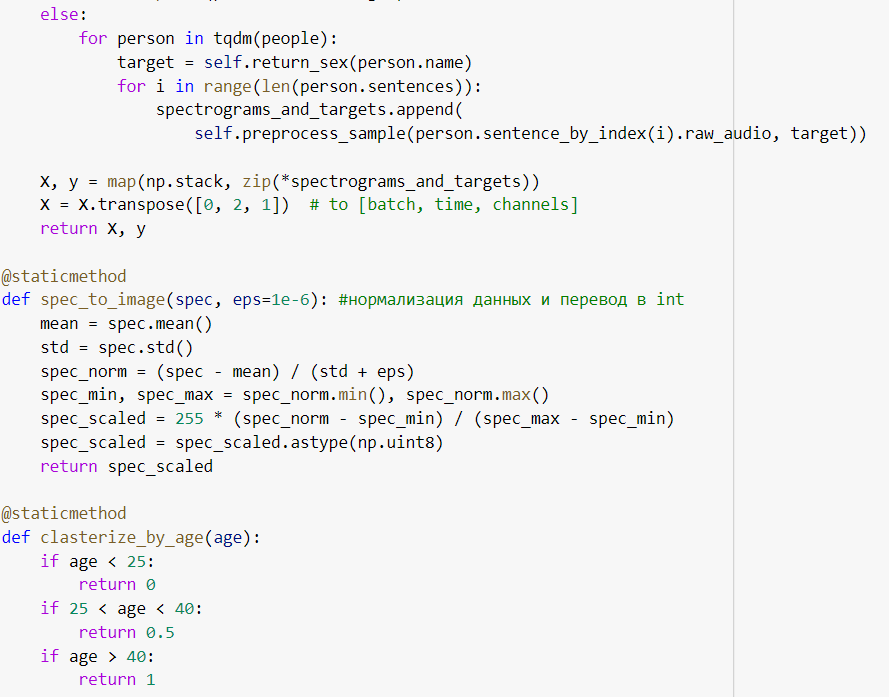


Рисунок 42 – Продолжение кода

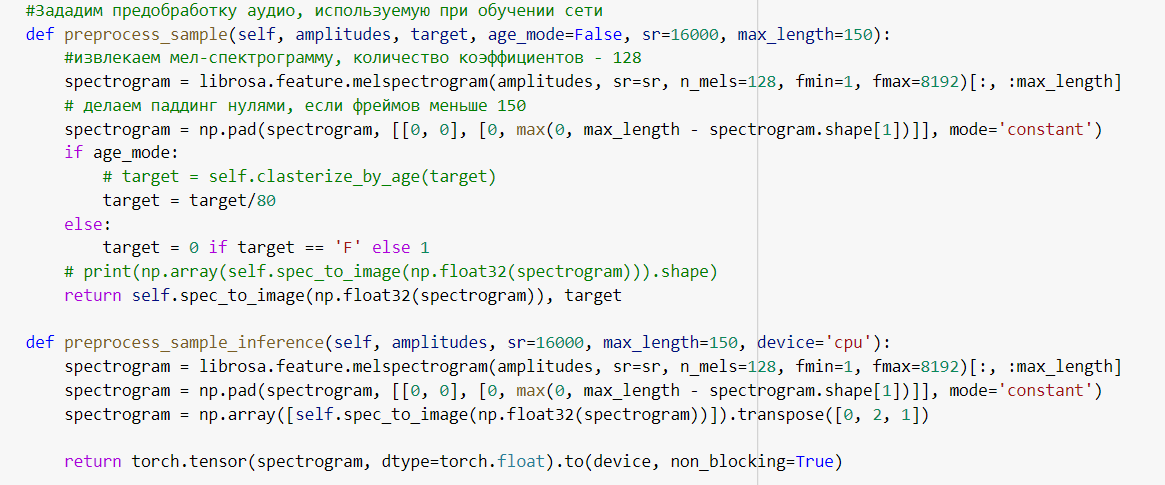


Рисунок 43 – Конец кода

Определим класс dataloader, который позволяет формировать батчи данных, подаваемые во время обучения нейросети.

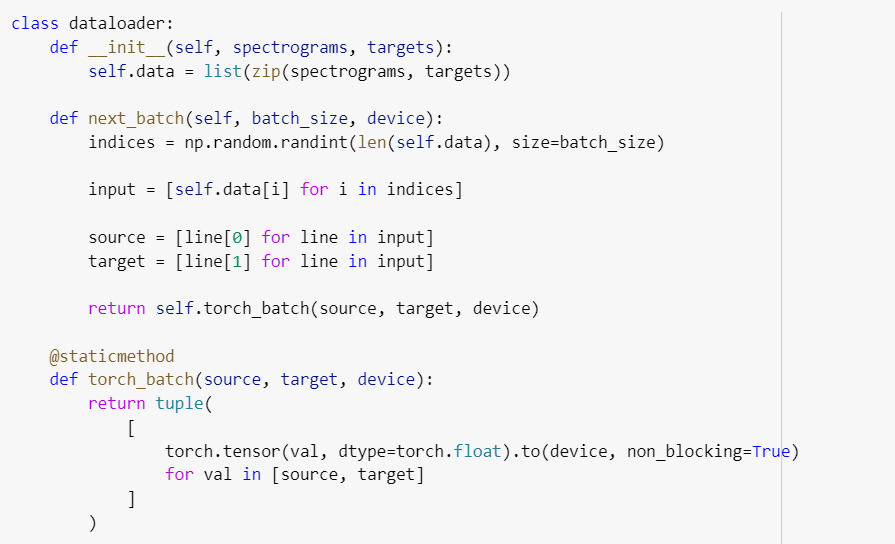


Рисунок 44 – Пример кода реализующий данный функционал

Зададим архитектуру обучаемой сети.

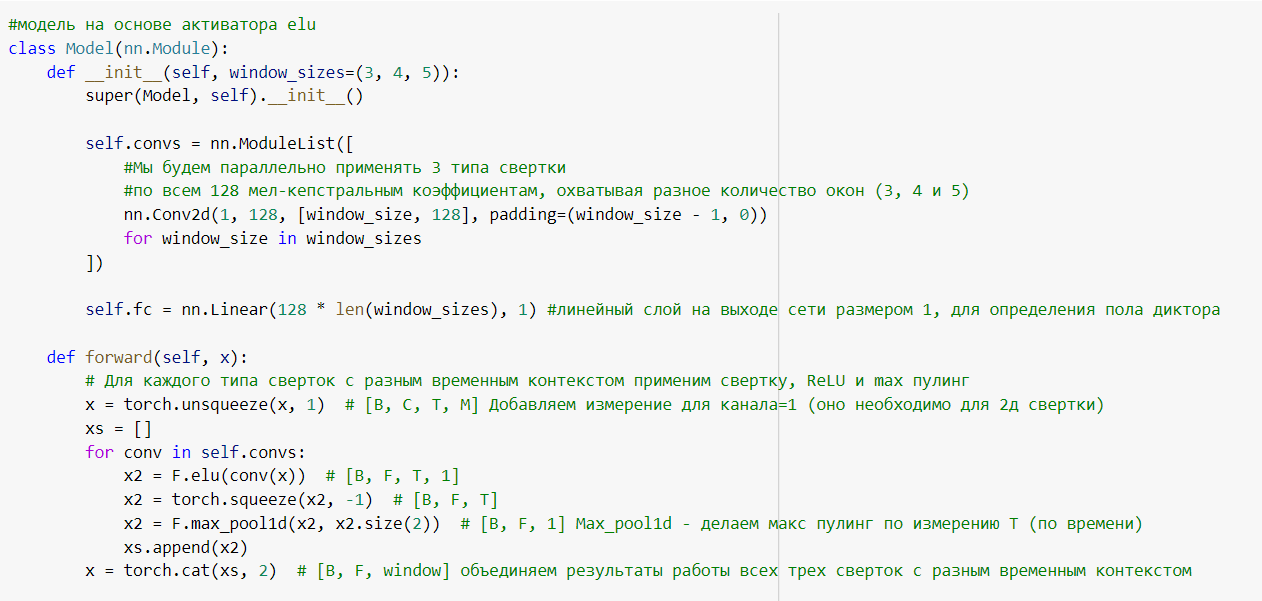




Рисунок 45 – Пример кода. Модель на основе активатора elu

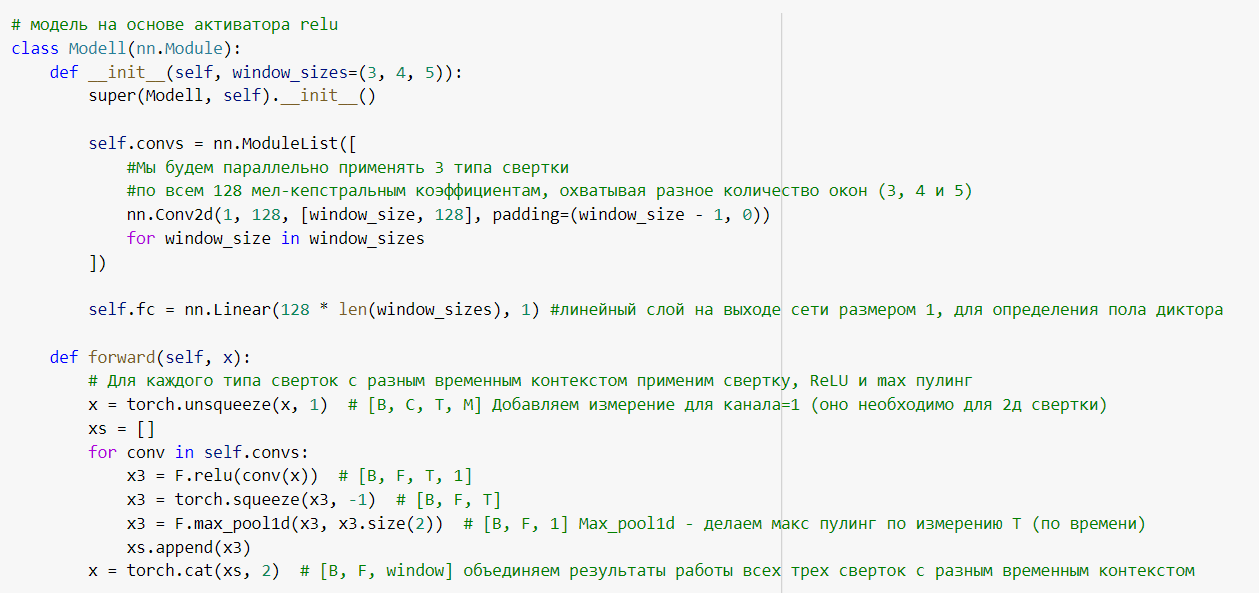




Рисунок 45 – Пример кода. Модель на основе активатора relu

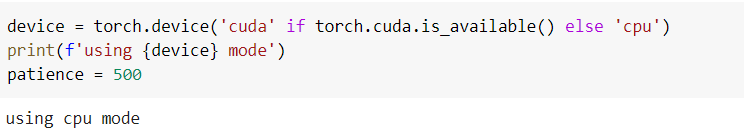
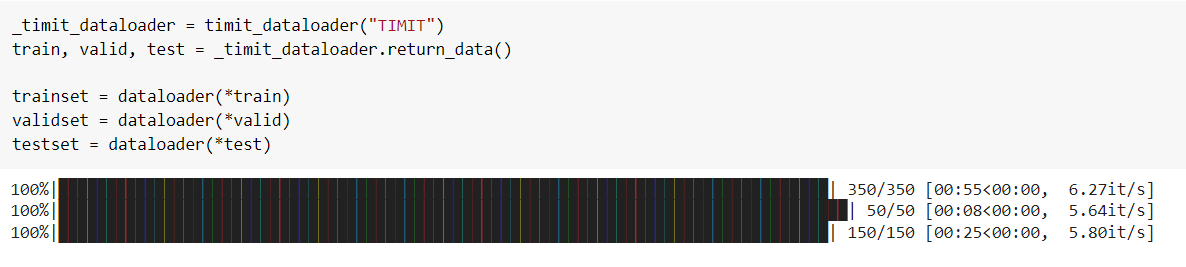
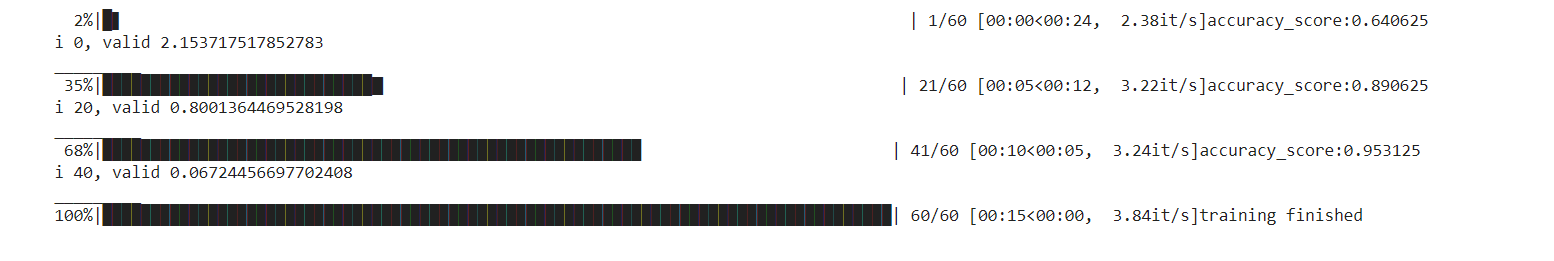


Рисунок 46 – Пример кода

Извлекаем мел-спектрограммы из всех аудиосигналов в наборе данных, подготовив обучающий, валидационный и тестовый наборы данных.



Обучаем нейронную сеть на основе активатора elu



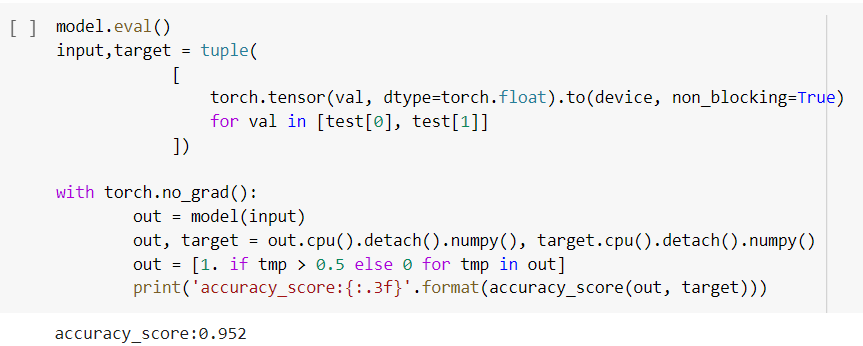
Обучаем нейронную сеть на основе активатора relu

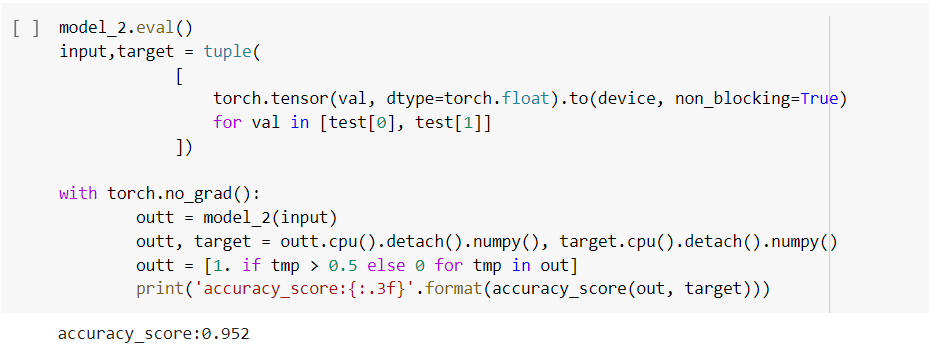


Рисунок 47 – Пример кода с результатом выполнения

* 1. **Тестирование модели**

После завершения обучения моделей, следует провести оценку их точности на обучающей и тестовой выборках. Это включает в себя вычисление метрик для оценки производительности обученных моделей. Для этого необходимо получить прогнозные вероятности на тестовом наборе данных.Начало формы



Рисунок 48 – Пример кода с получением вероятностей

Зададим функцию, позволяющую рассчитать равную ошибку 1-го и 2-го рода (equal error rate, EER).

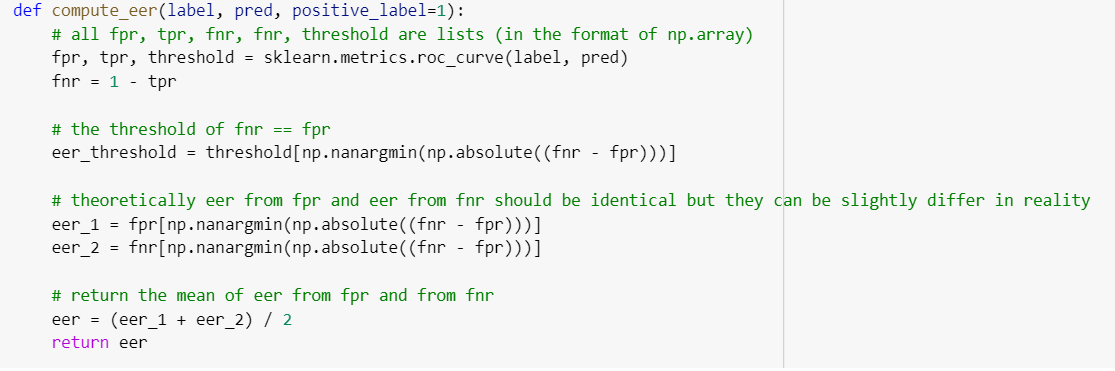


Рисунок 49 – Пример кода задаваемой функции

Давайте построим ROC-кривую и DET-кривую для оценки производительности наших моделей.

ROC-кривая (Receiver Operating Characteristic) представляет собой графическую характеристику качества бинарного классификатора. Она отражает зависимость доли верных положительных классификаций от доли ложных положительных классификаций при изменении порога решающего правила. Это позволяет нам оценить, насколько хорошо модель разделяет положительные и отрицательные классы.

DET-кривая (Detection Error Tradeoff) или кривая компромиссного определения ошибки показывает зависимость ошибки первого и второго рода бинарного классификатора от порога решающего правила. Эта кривая также помогает нам оценить производительность классификатора, особенно при важности баланса между ошибками 1-го и 2-го рода.

Построение и анализ таких кривых помогут нам лучше понять, как наши модели справляются с задачей классификации и выбрать оптимальный порог для принятия решений.

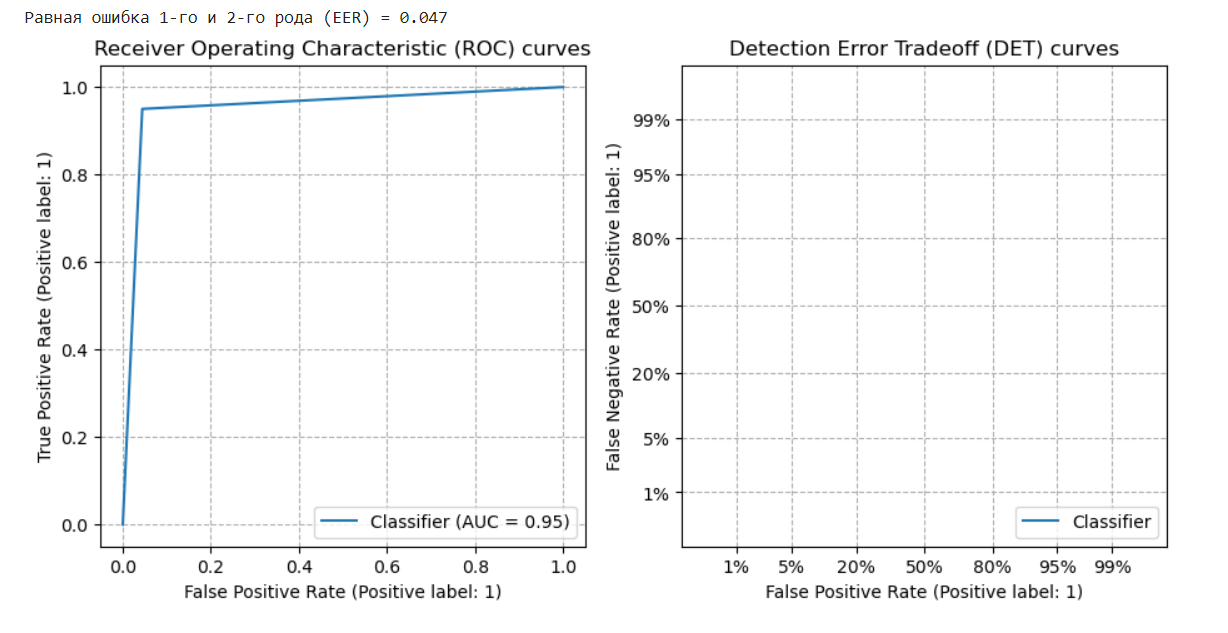
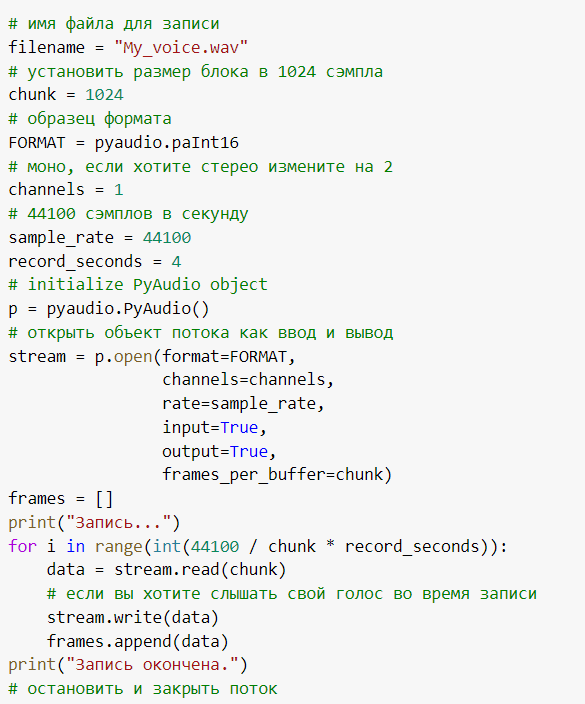


Рисунок 50 – График

* 1. **Проверка работы модели на собственном голосе**

Попробуем записать свой голос через микрофон и проверить, насколько хорошо работает модель с реальным голосом.



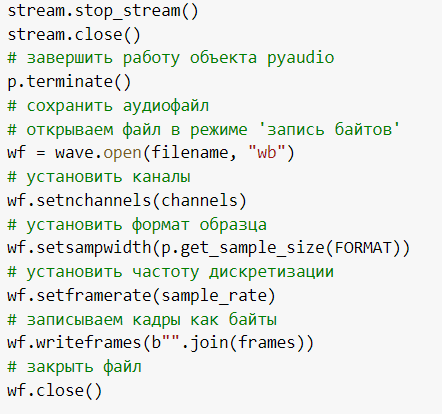
****

Рисунок 51 – Пример кода



Рисунок 52 – Пример кода

* 1. **Разработка приложения**

Завершающим этапом в нашем исследовательском проекте была разработка приложения на основе модели, которую мы обучили в ходе нашей работы. Для создания этого приложения мы использовали фреймворк Flask. Для начала, мы установили необходимую библиотеку Flask и определили маршруты с методами POST и GET.

Затем мы интегрировали нашу обученную глубокую сверточную нейронную сеть для классификации аудиоданных в приложение. Пользователи могут загружать аудиофайлы, и эти файлы проходят через нашу модель. В результате, мы получаем ответ о том, говорит ли на аудиофайле мужчина или женщина.

Мы также обнаружили, что разработка приложения включает в себя две основные составляющие: frontend и backend. Реализация backend прошла успешно, однако, у нас возникли некоторые сложности с frontend. В связи с ограниченным временем и ресурсами, которые мы имели в рамках курса, мы не смогли подробно изучить эту тему. Надеемся, что несмотря на возникшие сложности, нам удалось успешно решить все проблемы и создать работающее приложение.

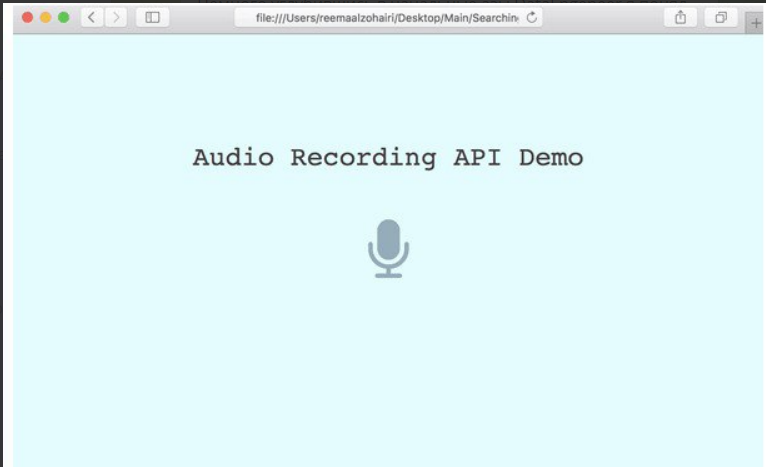


Рисунок 53 – Скриншоты нашей работы над приложением



Рисунок 54 – Скриншот не критичной ошибки

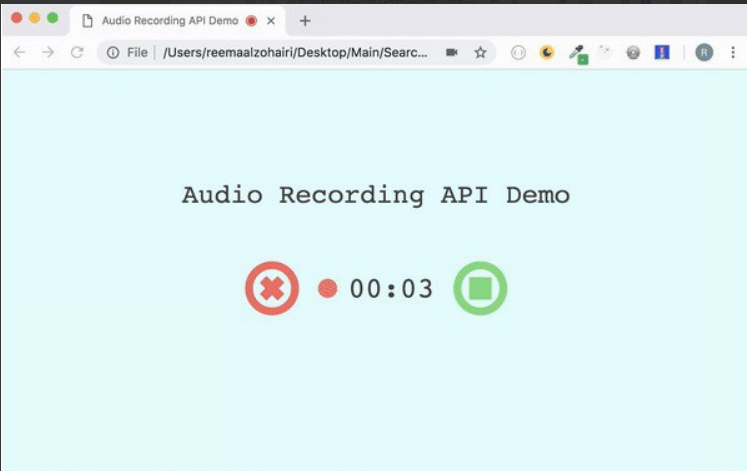


Рисунок 55 – Скриншот успешного начала записи

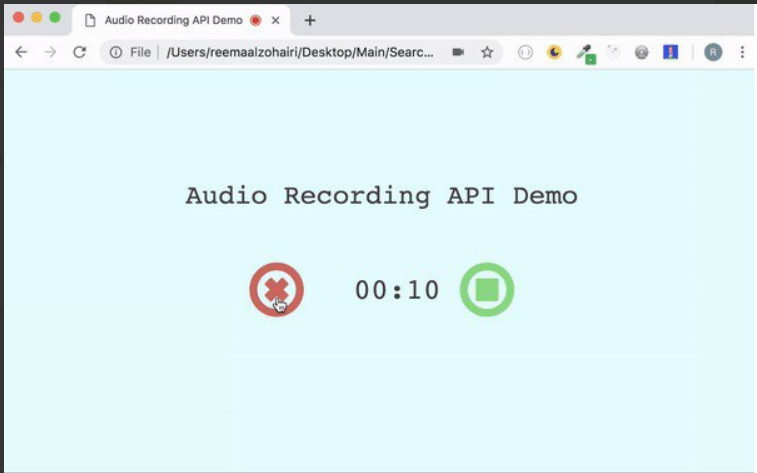


Рисунок 56 – Скриншот продолжения записи голоса

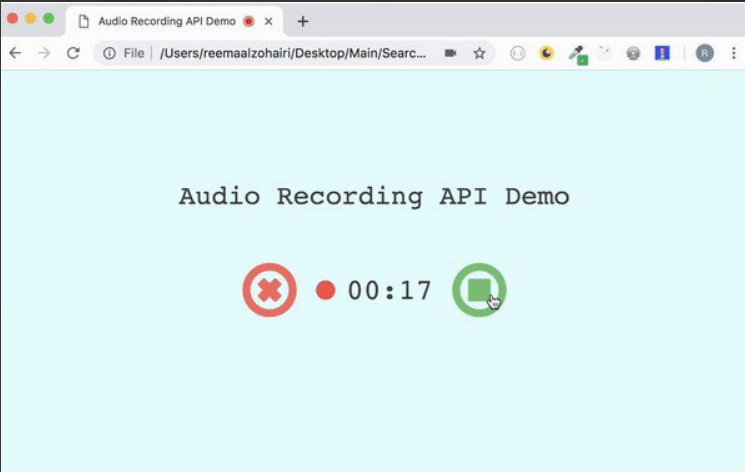


Рисунок 57 – Скриншот отсутствия ошибок по окончанию записи



Рисунок 58 – Пример кода

**ЗАКЛЮЧЕНИЕ**

Главной целью нашей исследовательской работы было разработать модель, способную с высокой точностью классифицировать аудиоданные по гендерному признаку.

Мы успешно достигли этой цели, применив глубокую сверточную нейронную сеть. Наша модель бинарной классификации аудиоданных оказалась очень эффективной и дает точность на уровне 95%. Это означает, что мы можем с высокой вероятностью определить пол человека по его голосу.

Мы надеемся, что данная модель будет внедрена и использована банками и компаниями, у которых есть контактные центры. Это позволит снизить случаи мошенничества при подтверждении финансовых операций через контактные центры, так как модель сможет выявлять несоответствия между голосом клиента и его учетными данными, что повысит безопасность финансовых операций.

**СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. Alex Maszański. Метод k-ближайших соседей (k-nearest neighbour): – Режим доступа: <https://proglib.io/p/metod-k-blizhayshih-sosedey-k-nearest-neighbour-2021-07-19>. (дата обращения: 01.09.2023)
2. Andre Ye. 5 алгоритмов регрессии в машинном обучении, о которых вам следует знать: – Режим доступа: <https://habr.com/ru/company/vk/blog/513842/> (дата обращения: 01.09.2023).
3. Белов, Е. Б. Основы информационной безопасности: учебное пособие / Е. Б. Белов, В. П. Лось – М. : Горячая линия – Телеком, 2006. – 544 с.
4. Бизли Д. Python. Подробный справочник: учебное пособие. – Пер. с англ. – СПб.: Символ-Плюс, 2010. – 864 с., ил.
5. Гафаров, Ф.М., Галимянов А.Ф. Искусственные нейронные сети и приложения: учеб. пособие /Ф.М. Гафаров, Галимянов А.Ф. – Казань: Издательство Казанского университета, 2018. – 121 с.
6. Грас Д. Data Science. Наука о данных с нуля: Пер. с англ. - 2-е изд., перераб. и доп. - СПб.: БХВ-Петербурr, 2021. - 416 с.: ил.
7. Головко В.А. От многослойных персептронов к нейронным сетям глубокого доверия: парадигмы обучения и применения // Лекции по нейроинформатике. – М., 2017. – С. 47-84.
8. Devpractice Team. Python. Визуализация данных. Matplotlib. Seaborn. Mayavi. - devpractice.ru. 2020. - 412 с.: ил.
9. Джулли, Пал: Библиотека Keras - инструмент глубокого обучения / пер. с англ. А. А. Слинкин.- ДМК Пресс, 2017. – 249 с.
10. Документация по библиотеке keras: – Режим доступа: <https://keras.io/api/>. (дата обращения: : 01.09.2023).
11. Документация по библиотеке librosa: – Режим доступа:  [https://librosa.org/](%20https://librosa.org/) . (дата обращения: : 01.09.2023).
12. Документация по библиотеке matplotlib: – Режим доступа: <https://matplotlib.org/stable/users/index.html>. (дата обращения: : 01.09.2023)
13. Документация по библиотеке numpy: – Режим доступа: <https://numpy.org/doc/1.22/user/index.html#user>. (дата обращения: : 01.09.2023).
14. Документация по библиотеке pandas: – Режим доступа: <https://pandas.pydata.org/docs/user_guide/index.html#user-guide>. (дата обращения: : 01.09.2023).
15. Документация по библиотеке scikit-learn: – Режим доступа: <https://scikit-learn.org/stable/user_guide.html>. (дата обращения: : 01.09.2023).
16. Документация по библиотеке seaborn: – Режим доступа: <https://seaborn.pydata.org/tutorial.html>. (дата обращения: : 01.09.2023).
17. Документация по библиотеке PyTorch: – Режим доступа: https://pytorch.org/docs/stable/index.html (дата обращения: : 01.09.2023).
18. Документация по языку программирования python: – Режим доступа:  <https://docs.python.org/3.8/index.html>. (дата обращения: : 01.09.2023).
19. Иванов Д.А., Ситников А.И., Шляпин С.Д – Композиционные материалы: учебное пособие для вузов, 2019. 13 с.
20. Краткий обзор алгоритма машинного обучения Метод Опорных Векторов (SVM) – Режим доступа: <https://habr.com/ru/post/428503/> (дата обращения: : 01.09.2023)
21. Курс «Нейронные сети и компьютерное зрение» – Режим доступа: <https://stepik.org/course/50352/> (дата обращения: : 01.09.2023)
22. Миронов А.А. Машинное обучение часть I ст.9 – Режим доступа: <http://is.ifmo.ru/verification/machine-learning-mironov.pdf>. (дата обращения: : 01.09.2023)
23. Николенко С., Кадурин А., Архангельская Е. Глубокое обучение.Погружение в мир нейронных сетей. – СПб.: Питер, 2019. – 480 с.
24. Плас Дж. Вандер, Python для сложных задач: наука о данных и машинное обучение. Санкт-Петербург: Питер, 2018, 576 с.
25. Роббинс, Дженнифер. HTML5: карманный справочник, 5-е издание.: Пер. с англ. - М.: ООО «И.Д. Вильямс»: 2015. - 192 с.: ил.
26. Руководство по быстрому старту в flask: – Режим доступа: <https://flask-russian-docs.readthedocs.io/ru/latest/quickstart.html>. (дата обращения: : 01.09.2023)
27. Силен Дэви, Мейсман Арно, Али Мохамед. Основы Data Science и Big Data. Python и наука о данных. – СПб.: Питер, 2017. – 336 с.: ил.
28. Скиена, Стивен С. С42 Наука о данных: учебный курс.: Пер. с англ. - СПб.: ООО "Диалектика", 2020. - 544 с.
29. Траск Э. Грокаем глубокое обучение. – СПб.: Питер, 2019. – 352с.
30. Тропченко А.Ю., Тропченко А.А. Методы сжатия изображений, аудиосигналов и видео. Учебное пособие по дисциплине «Теоретическая информатика» - Санкт-Петербург: 2009. - 108 с.