ГАОУ СО “СамЛИТ (Базовая школа РАН)”

**ИНДИВИДУАЛЬНЫЙ ПРОЕКТ**

**Разработка системы распознавания эмоций в голосе человека.**

Работу выполнил:

Пестровский Евгений Михайлович,

ученик 10 “А”

Руководитель:

Марков Алексей Анатольевич

Учитель информатики:

Анисимова Светлана Анатольевна

Самара, 2023

**Содержание**

[**Введение** 3](#_Toc133314722)

[**Цель работы**: 3](#_Toc133314723)

[**Задачи:** 3](#_Toc133314724)

[**Вступление:** 3](#_Toc133314725)

[**Поиск информации о существующих информационных системах обработки голоса** 4](#_Toc133314726)

[**Поиск подходящей среды программирования** 6](#_Toc133314727)

[**Поиск подходящих библиотек и фреймворков для работы с аудиофайлами, нейронными сетями и графическим интерфейсом** 6](#_Toc133314728)

[**Поиск подходящих наборов данных голоса** 7](#_Toc133314729)

[**Процесс обучения нейронной сети** 8](#_Toc133314730)

[**Создание графического интерфейса для взаимодействия с пользователем** 9](#_Toc133314731)

[**Заключение** 10](#_Toc133314732)

[**Список использованной литературы** 10](#_Toc133314733)

# **Введение**

## **Цель работы**:

Создать систему распознавания эмоций в голосе человека, используя алгоритмы машинного обучения.

## **Задачи:**

1. Поиск информации о существующих информационных системах обработки голоса
2. Поиск подходящей среды программирования
3. Поиск подходящих библиотек и фреймворков для работы с аудиофайлами, нейронными сетями и графическим интерфейсом
4. Поиск подходящих наборов данных голоса
5. Описание процесса обучения нейронной сети и обучение нейронной сети
6. Создание графического интерфейса для взаимодействия с пользователем
7. Процесс подготовки демонстрационных материалов
8. Защита проекта

## **Вступление:**

В последние годы разработка систем распознавания эмоций в голосе человека стала актуальной задачей для многих научных и технологических команд. Эта тема имеет огромное значение для понимания и анализа эмоционального состояния человека в различных сферах жизни.

Распознавание эмоций в голосе может быть полезным в многих областях, включая медицину, психологию, маркетинг, образование и бизнес. Например, в колл-центрах и других местах обслуживания клиентов, системы распознавания эмоций могут помочь идентифицировать эмоциональное состояние клиентов и повысить качество обслуживания, предлагая соответствующие решения.

Также, системы распознавания эмоций могут использоваться в медицине для диагностики психических расстройств и других заболеваний, связанных с эмоциональным состоянием человека. В психологии они могут помочь исследователям более точно анализировать эмоциональные реакции и эмоциональную регуляцию.

В данном проекте я буду разрабатывать систему распознавания эмоций в голосе человека с помощью методов машинного обучения. Она будет способна определять различные эмоции, такие как радость, грусть, злость, страх и т. д. Моя цель - создать инструмент, который будет полезен в различных областях и поможет улучшить качество обслуживания клиентов, повысить эффективность лечения пациентов и улучшить жизнь людей в целом.

# **Поиск информации о существующих информационных системах обработки голоса**

Существует множество систем распознавания эмоций в голосе, которые используют различные методы и алгоритмы:

1. Affectiva - это платформа для распознавания эмоций в реальном времени, использующая нейросетевые алгоритмы для анализа эмоций по видео- и аудиоданным.
2. EmoVoice - это система распознавания эмоций в голосе, разработанная исследователями университета Хельсинки. Она использует алгоритмы машинного обучения для анализа различных параметров голоса, таких как тембр, интонация, скорость и громкость, для определения эмоционального состояния говорящего.
3. Beyond Verbal - это система анализа голоса, которая использует десятки тысяч параметров голоса для определения эмоционального состояния говорящего. Она также может определять другие параметры, такие как возраст, пол и физическое состояние.
4. Nemesysco - это компания, специализирующаяся на разработке систем распознавания эмоций в голосе. Они используют методы анализа сигналов для измерения физиологических параметров голоса, таких как частота дыхания, сердцебиение и т.д., для определения эмоционального состояния говорящего.
5. IBM Watson - это система искусственного интеллекта, которая включает в себя функции распознавания эмоций в голосе. Она использует методы машинного обучения для анализа параметров голоса и определения эмоционального состояния говорящего.
6. Speech Technology Center (STC) - это российская компания, которая занимается разработкой систем распознавания речи и анализа эмоций в голосе. Их технологии используются в таких сферах, как банковское дело, государственное управление, медицина и др.
7. Cognitive Pilot - это российская компания, которая разрабатывает автоматизированные системы управления транспортом, в том числе системы распознавания эмоций водителя. Эта система использует алгоритмы машинного обучения для анализа голоса и определения эмоционального состояния водителя, что позволяет предотвращать возможные аварии.
8. i-Free - это российская компания, которая занимается разработкой мобильных приложений и сервисов. Они также разработали систему распознавания эмоций в голосе под названием "EmoClick", которая может использоваться для анализа эмоционального состояния пользователей и улучшения взаимодействия с ними.
9. Neurodata Lab - это российская компания, которая занимается разработкой систем распознавания эмоций в голосе и на лице. Их технологии используются в различных сферах, таких как маркетинг, реклама, образование и др.

Конечно, это лишь несколько примеров из множества существующих систем распознавания эмоций в голосе, и каждая из них имеет свои особенности и преимущества.

# **Поиск подходящей среды программирования**

При разработке был проведен поиск подходящей среды программирования. Для создания и обучения модели нейросети был использован Jupyter Notebook - это интерактивная среда программирования, которая позволяет выполнять код на различных языках программирования, включая Python, используя веб-браузер. Jupyter Notebook удобен для проведения исследований, создания прототипов моделей и отладки кода.

Для создания графического интерфейса, с которым будет работать пользователь, был выбран PyCharm - это интегрированная среда разработки на языке программирования Python, которая позволяет создавать приложения с графическим интерфейсом. PyCharm предоставляет различные инструменты для удобной разработки, отладки и тестирования приложений, включая автодополнение кода, систему контроля версий, подсветку синтаксиса и многое другое.

Таким образом, использование Jupyter Notebook и PyCharm позволило эффективно разрабатывать и обучать модель нейросети, а также создать удобный и интуитивно понятный графический интерфейс для пользователей.

# **Поиск подходящих библиотек и фреймворков для работы с аудиофайлами, нейронными сетями и графическим интерфейсом**

Для работы с аудиофайлами была использована библиотека librosa, которая позволяет извлекать различные признаки из аудиофайлов, такие как спектрограммы, мел-частотные кепстральные коэффициенты (MFCC), и другие.

Для создания и обучения модели нейросети был использован фреймворк Scikit-learn(sklearn), который предоставляет широкий спектр инструментов для реализации различных алгоритмов машинного обучения, включая нейронные сети. Sklearn удобен для создания и обучения моделей и достаточно прост в использовании, а также для оценки их эффективности.

Для создания графического интерфейса был выбран фреймворк PyQt5, являющийся кроссплатформенным инструментом для создания графических приложений на Python. PyQt5 предоставляет множество инструментов для разработки пользовательского интерфейса, таких как кнопки, поля ввода, окна и другие элементы управления.

Также для работы с данными были использованы библиотеки numpy и matplotlib. Numpy позволяет выполнять вычисления с многомерными массивами, что необходимо при работе с табличными данными. Matplotlib, в свою очередь, позволяет строить различные графики и диаграммы для визуализации данных.

Таким образом, выбор библиотек и фреймворков основывался на их удобстве использования, наличии необходимых инструментов для решения поставленных задач, а также на широкой поддержке сообщества и наличии достаточной документации.

# **Поиск подходящих наборов данных голоса**

Для проекта был выбран датасет Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS). Этот датасет содержит только аудиофайлы речи (16 бит, 48 кГц .wav) и включает в себя 1440 файлов: 60 проб на актера x 24 актера = 1440. В датасете представлены 24 профессиональных актера (12 женщин и 12 мужчин), произносящих два выражения с нейтральным североамериканским акцентом. Эмоциональные выражения включают спокойствие, радость, грусть, злость, страх, удивление и отвращение, каждое выражение произносится в двух уровнях эмоциональной интенсивности (нормальный, сильный), с дополнительным нейтральным выражением.

# **Процесс обучения нейронной сети**

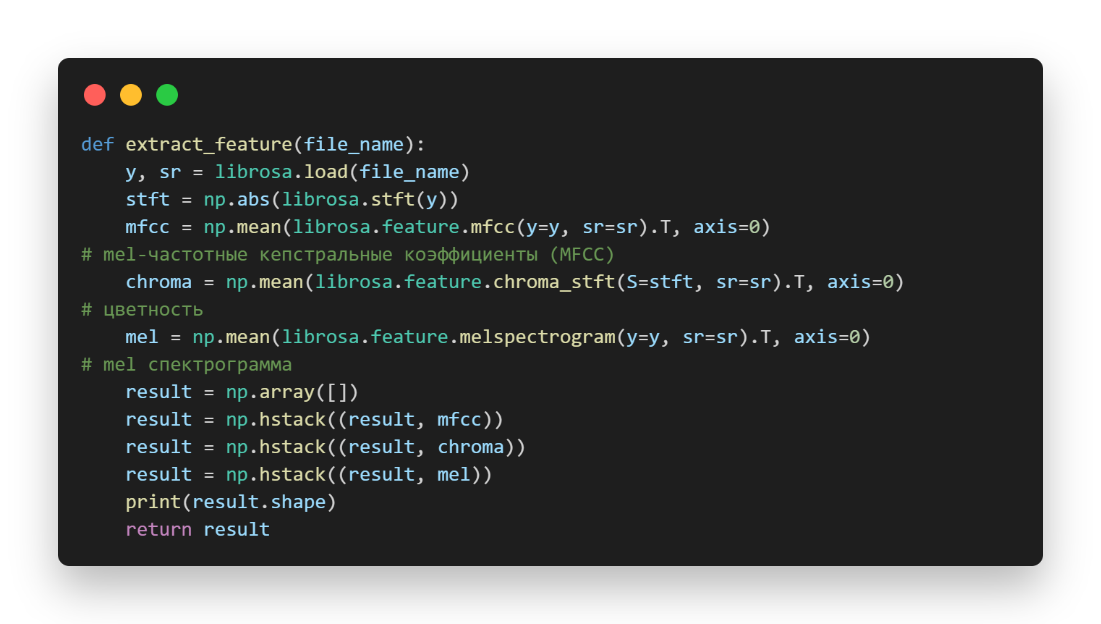
Для создания и обучения нейронной сети был выбран полносвязный перцептрон MLPClassifier в библиотеке sklearn.

Были использованы следующие настройки:



Все признаки извлекались из аудиофайлов с помощью библиотеки librosa.

Извлечение признаков из аудиофайла:



В начале обучения нейросеть пыталась распознать все восемь эмоций, предоставляемых датасетом, однако качество распознавания было довольно низким - примерно 41% на тестовых данных. Поэтому было принято решение упростить задачу и оставить только три основных эмоции: радость, спокойствие и гнев.

В результате этого упрощения задачи качество распознавания улучшилось и составило 83% на тестовых данных. Однако стоит понимать, что на реальных данных качество распознавания может сильно упасть.

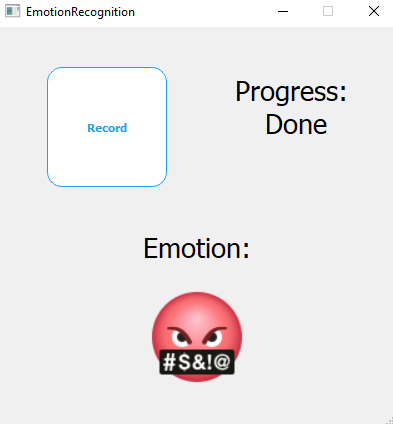
# **Создание графического интерфейса для взаимодействия с пользователем**

Для создания графического интерфейса была выбрана библиотека PyQt5, которая предоставляет широкие возможности для создания функциональных и эстетически привлекательных приложений с графическим интерфейсом. Одним из основных инструментов для разработки интерфейса была использована утилита Qt Designer, которая позволяет создавать интерфейсы путем перетаскивания элементов на форму, а затем экспортировать их в код на Python.

В интерфейсе были предусмотрены элементы для записи аудиофайла, а также для распознавания эмоций в голосе. Запись голоса осуществлялась с помощью класса QAudioRecorder, который предоставляет функциональность для записи звука с микрофона компьютера. После записи звука он сохраняется в файл формата .wav.

После того, как была записана речь, она автоматически передается в нейронную сеть для анализа. Результат распознавания выводится на экран в виде текстовой метки, которая содержит картинку-эмоджи, соответствующую наиболее вероятной эмоции.

В целом, графический интерфейс был разработан с учетом удобства и интуитивной понятности для пользователя. Он позволяет быстро записывать и анализировать голосовые сообщения на предмет выражения эмоций в голосе.



# **Заключение**

В процессе разработки моего проекта по разработке системы распознавания эмоций в голосе человека я получил много ценного опыта в области машинного обучения и анализа данных. Было доказано, что такая система имеет большой потенциал для применения в различных сферах жизни, таких как медицина, психология, маркетинг, образование и бизнес.

Однако, следует отметить, что дальнейшее совершенствование системы распознавания эмоций в голосе человека требует дополнительных исследований и разработок. Например, необходимо учитывать культурные и языковые особенности разных стран и регионов для повышения точности распознавания эмоций.

В целом, проект имеет большое значение для улучшения качества жизни людей. Я надеюсь, что мой инструмент будет полезен для решения многих практических задач и сможет помочь людям в их повседневной жизни.

# **Список использованной литературы**

1. <https://www.kaggle.com/datasets/uwrfkaggler/ravdess-emotional-speech-audio>
2. <https://data-flair.training/blogs/python-mini-project-speech-emotion-recognition/>
3. <https://librosa.org/doc/latest/index.html>
4. <https://scikit-learn.org/stable/>
5. <https://numpy.org/doc/>
6. <https://doc.qt.io/qtforpython-6/>
7. <https://docs.jupyter.org/en/latest/>
8. <https://www.jetbrains.com/pycharm/>