VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI – 590018, Karnataka

INTERNSHIP REPORT

ON

"Voice Classification using ML"

Submitted in partial fulfilment for the award of degree(21INT49)

**BACHELOR OF ENGINEERING IN
COMPUTER SCIENCE AND
ENGINEERING (DATA SCIENCE)**

Submitted by:

CHARITHA DEEPAK

4AD21CD011

Conducted at
Compsoft Technologies

ATME COLLEGE OF ENGINEERING

Department of Computer science and engineering-data science

Accredited by NBA, New Delhi
Bannur road, Mysuru, Karnataka 570028

ATME COLLEGE OF ENGINEERING

Department of Computer science and engineering-data science

Accredited by NBA, New Delhi

Bannur road, Mysuru, Karnataka 570028

CERTIFICATE

This is to certify that the Internship titled **“Voice Classification using ML”** carried out by **Ms. CHARITHA DEEPAK,** a bonafide student of ATME college of engineering, in partial fulfillment for the award of **Bachelor of Engineering**, in **Computer science and engineering-Data science** under Visvesvaraya Technological University,Belagavi, during the year 2022-2023. It is certified that all corrections/suggestions indicated have been incorporated in the report.

The project report has been approved as it satisfies the academic requirements in respect of Internship prescribed for the course Internship / Professional Practice (21INT49)

Signature of Guide **Signature of HOD** **Signature of Principal**

External Viva:

Name of the Examiner Signature with Date

1)__________________

____

2)__________________

____

DECLARATION

I, **Charitha Deepak** third year student of Computer science and engineering-Data science, ATME college of engineering - 560 028, declare that the Internship has been successfully completed, in **COMPSOFT TECHNOLOGIES**. This report is submitted in partial fulfillment of the requirements for award of Bachelor Degree in CSE-data science, during the academic year 2022-2023.

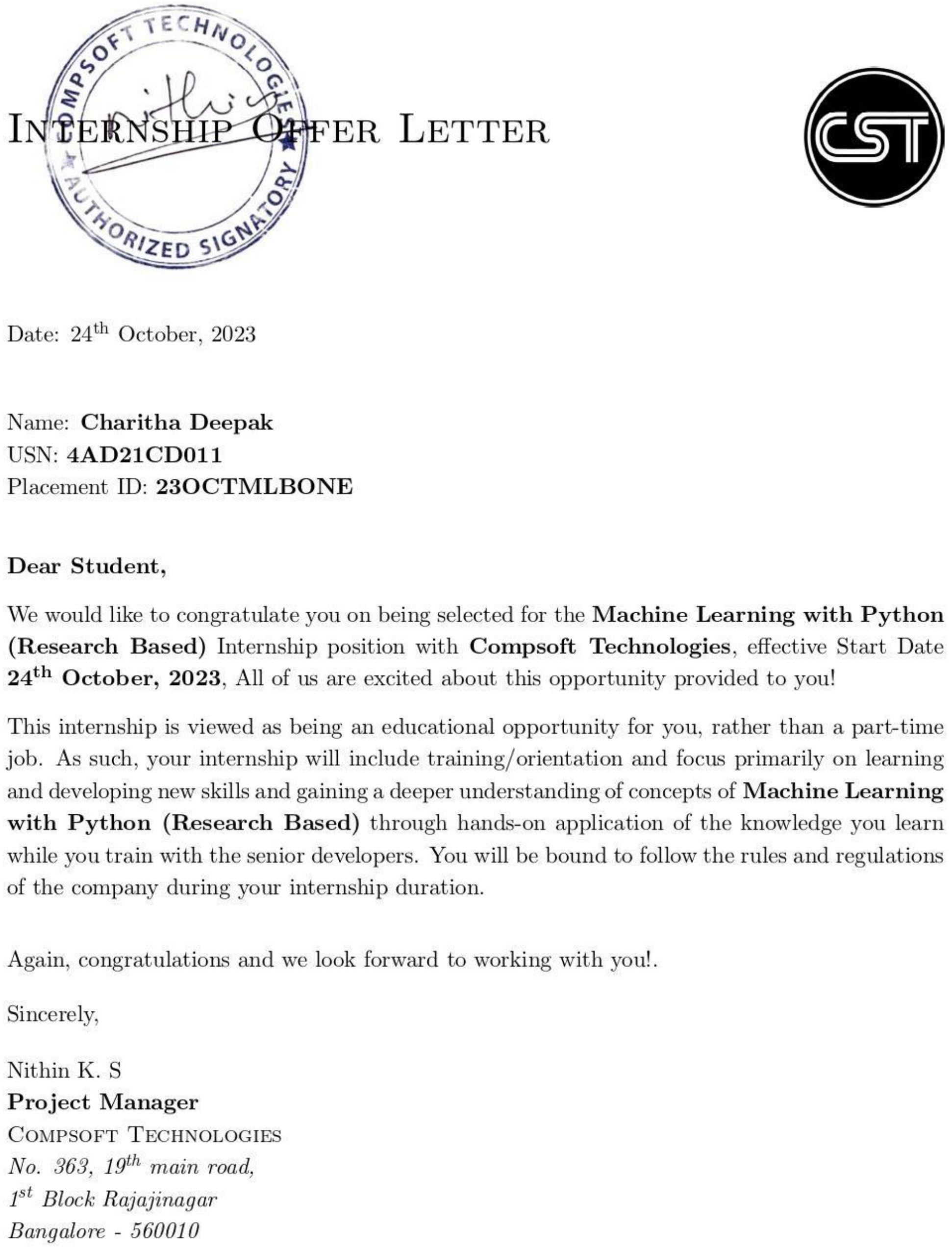
Date :30/11/2023 :

Place : Mysuru

USN :4AD21CD011

NAME : CHARITHA DEEPAK

OFFER LETTER



COMPSOFT TECHNOLOGIES ★ AUTHORIZED SIGNATORY

Internship Offer Letter

Date: 24th October, 2023

Name: **Charitha Deepak**
USN: **4AD21CD011**
Placement ID: **23OCTMLBONE**

Dear Student,

We would like to congratulate you on being selected for the **Machine Learning with Python (Research Based)** Internship position with **Compsoft Technologies**, effective Start Date **24th October, 2023**, All of us are excited about this opportunity provided to you!

This internship is viewed as being an educational opportunity for you, rather than a part-time job. As such, your internship will include training/orientation and focus primarily on learning and developing new skills and gaining a deeper understanding of concepts of **Machine Learning with Python (Research Based)** through hands-on application of the knowledge you learn while you train with the senior developers. You will be bound to follow the rules and regulations of the company during your internship duration.

Again, congratulations and we look forward to working with you!.

Sincerely,

Nithin K. S
Project Manager
Compsoft Technologies
No. 363, 19th main road,
1st Block Rajajinagar
Bangalore - 560010

ACKNOWLEDGEMENT

This Internship is a result of accumulated guidance, direction and support of several important persons. We take this opportunity to express our gratitude to all who have helped us to complete the Internship.

We express our sincere thanks to our Principal, for providing us adequate facilities to undertake this Internship.

We would like to thank our Head of Dept –computer science and engineering (Data science), for providing us an opportunity to carry out Internship and for his valuable guidance and support.

We would like to thank our (Lab assistant) Software Services for guiding us during the period of internship.

We express our deep and profound gratitude to our guide, Prof Neethi M.V, Assistant Prof, for her keen interest and encouragement at every step in completing the Internship.

We would like to thank all the faculty members of our department for the support extended during the course of Internship.

We would like to thank the non-teaching members of our dept, for helping us during the Internship.

Last but not the least, we would like to thank our parents and friends without whose constant help, the completion of Internship would have not been possible.

CHARITHA DEEPAK
4AD21CD011

ABSTRACT

This project delves into the realm of voice classification through the lens of machine learning (ML). The primary objective is to develop a robust system capable of accurately categorizing diverse voice samples. The process begins with thorough data preprocessing to ensure data quality and consistency. Feature extraction techniques are then applied to capture relevant aspects of the audio data.

Various ML models, such as support vector machines, neural networks, and ensemble methods, are implemented and fine-tuned to achieve optimal performance. The project also explores the impact of different feature sets on classification accuracy. Rigorous evaluation metrics are employed to assess the models' effectiveness in handling various voice patterns and nuances.

Additionally, real-world applications and implications of voice classification, such as in speech recognition systems, security protocols, and personalized user experiences, are discussed. The findings of this project contribute to the broader field of voice analysis and have potential implications for improving the functionality and accuracy of voice-related technologies.

Table of Contents

CHAPTER 1

COMPANY PROFILE

1. COMPANY PROFILE

A Brief History of Compsoft Technologies

Compsoft Technologies, was incorporated with a goal To provide high quality and optimal Technological Solutions to business requirements of our clients". Every business is a different and has a unique business model and so are the technological requirements. They understand this and hence the solutions provided to these requirements are different as well. They focus on clients requirements and provide them with tailor made technological solutions. They also understand that Reach of their Product to its targeted market or the automation of the existing process into e-client and simple process are the key features that our clients desire from Technological Solution they are looking for and these are the features that we focus on while designing the solutions for their clients.

Sarvamoola Software Services. is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever increasing automation requirements, Sarvamoola Software Services. specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting clients requirements.

Compsoft Technologies, strive to be the front runner in creativity and innovation in software development through their well-researched expertise and establish it as an out of the box software development company in Bangalore, India. As a software development company, they translate this software development expertise into value for their customers through their professional solutions.

They understand that the best desired output can be achieved only by understanding the clients demand better. Compsoft Technologies work with their clients and help them to defiine their exact solution requirement. Sometimes even they wonder that they have completely redefined their solution or new application requirement during the brainstorming session, and here they position themselves as an IT solutions consulting group comprising of high caliber consultants.

They believe that Technology when used properly can help any business to scale and achieve new heights of success. It helps Improve its efficiency, profitability, reliability; to put itin one sentence " Technology helps you to Delight your Customers" and that is what we want to achieve.

CHAPTER 2

ABOUT THE COMPANY

2. ABOUT THE COMPANY

Compsoft Technologies is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever increasing automation requirements, Compsoft Technologies specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting clients requirements. The organization where they have a right mix of professionals as a stakeholders to help us serve our clients with best of our capability and with at par industry standards. They have young, enthusiastic, passionate and creative Professionals to develop technological innovations in the field of Mobile technologies, Web applications as well as Business and Enterprise solution. Motto of our organization is to "Collaborate with our clients to provide them with best Technological solution hence creating Good Present and Better Future for our client which will bring a cascading a positive effect in their business shape as well". Providing a Complete suite of technical solutions is not just our tag line, it is Our Vision for Our Clients and for Us, We strive hard to achieve it.

Products of Compsoft Technologies.

Android Apps

It is the process by which new applications are created for devices running the Android operating system. Applications are usually developed in Java (and/or Kotlin; or other such option) programming language using the Android software development kit (SDK), but other development environments are also available, some such as Kotlin support the exact same Android APIs (and bytecode), while others such as Go have restricted API access.

The Android software development kit includes a comprehensive set of development tools. These include a debugger, libraries, a handset emulator based on QEMU, documentation, sample code, and zutorials. Currently supported development platforms include computers running Linux (any modern desktop Linux distribution), Mac OS X 10.5.8 or later, and Windows 7 or later. As of March 2015, the SDK is not available on Android itself, but softwaredevelopment is possible by using specialized Android applications.

Web Application

It is a client–server computer program in which the client (including the user interface and client- side logic) runs in a web browser. Common web applications include web mail, online retail sales, online auctions, wikis, instant messaging services and many other functions. web applications use web documents written in a standard format such as HTML and JavaScript,which are supported by a variety of web browsers. Web applications can be considered as a specific variant of client–server software where the client software is downloaded to the client machine when visiting the relevant web page, using standard procedures such as HTTP. The Client web software updates may happen each time the web page is visited. During the session, the web browser interprets and displays the pages, and acts as the universal client for any web application. The use of web application frameworks can often reduce the number of errors in a program, both by making the code simpler, and by allowing one team to concentrate on the framework while another focuses on a specified use case. In applications which are exposed to constant hacking attempts on the Internet, security-related problems can be caused by errors in the program.

Frameworks can also promote the use of best practices such as GET after POST. There are some who view a web application as a two-tier architecture. This can be a “smart” client that performs all the work and queries a “dumb” server, or a “dumb” client that relies on a “smart” server. The client would handle the presentation tier, the server would have the database (storage tier), and the business logic (application tier) would be on one of them or on both. While this increases the scalability of the applications and separates the display and the database, it still doesn‟t allow for true specialization of layers, so most applications will outgrow this model. An emerging strategy for application software companies is to provide web access to software previously distributed as local applications. Depending on the type of application, it may require the development of an entirely different browser-based interface, or merely adapting an existing application to use different presentation technology. These programs allow the user to pay a monthly or yearly fee for use of a software application without having to install it on a local hard drive. A company which follows this strategy is known as an application service provider (ASP), and ASPs are currently receiving much attention in the software industry.

Security breaches on these kinds of applications are a major concern because it can involve both enterprise information and private customer data. Protecting these assets is an important part of any web application and there are some key operational areas that must be included in the development process. This includes processes for authentication, authorization, asset handling, input, and logging and auditing. Building security into the applications from the beginning can be more effective and less disruptive in the long run.

Web design

It is encompasses many different skills and disciplines in the production and maintenance of websites. The different areas of web design include web graphic design; interface design; authoring, including standardized code and proprietary software; user experience design; and search engine optimization. The term web design is normally used to describe the design process relating to the front-end (client side) design of a website including writing mark up. Web design partially overlaps web engineering in the broader scope of web development. Web designers are expected to have an awareness of usability and if their role involves creating mark up then they are also expected to be up to date with web accessibility guidelines. Web design partially overlaps web engineering in the broader scope of web development.

Departments and services offered

Compsoft Technologies plays an essential role as an institute, the level of education, development of student’s skills are based on their trainers. If you do not have a good mentor then you may lag in many things from others and that is why we at Compsoft Technologies gives you the facility of skilled employees so that you do not feel unsecured about the academics. Personality development and academic status are some of those things which lie on mentor’s hands. If you are trained well then you can do well in your future and knowing its importance of Compsoft Technologies always tries to give you the best.

They have a great team of skilled mentors who are always ready to direct their trainees in the best possible way they can and to ensure the skills of mentors we held many skill development programs as well so that each and every mentor can develop their own skills with the demands of the companies so that they can prepare a complete packaged trainee.

Services provided by Compsoft Technologies.

- Core Java and Advanced Java
- Web services and development
- Dot Net Framework
- Python
- Selenium Testing
- Conference / Event Management Service
- Academic Project Guidance
- On The Job Training
- Software Training

CHAPTER 3

INTRODUCTION

3. INTRODUCTION

Introduction to ML

Machine Learning (ML) is a field of artificial intelligence that empowers computers to learn and improve from experience without being explicitly programmed. Instead of relying on explicit programming instructions, ML algorithms use statistical patterns and computational models to make predictions or decisions based on data.

The core concept of ML is to enable machines to automatically learn and adapt from data, allowing them to generalize and make accurate predictions on new, unseen information. This learning process involves the identification of patterns, relationships, and trends within datasets, enabling machines to make informed decisions or predictions in various domains, from image recognition and natural language processing to recommendation systems and autonomous vehicles.

There are three main types of machine learning: supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, models are trained on labeled data, where the algorithm learns to map input data to corresponding output labels. Unsupervised learning involves exploring and identifying patterns in unlabeled data, while reinforcement learning focuses on training models through interactions with an environment, learning from trial and error.

Machine Learning plays a crucial role in today's technology landscape, powering a wide range of applications that impact our daily lives, from personalized recommendations on streaming platforms to medical diagnosis and autonomous systems. As the field continues to evolve, machine learning promises to bring forth new capabilities and innovations, shaping the future of artificial intelligence.

Types of Machine Learning:

1. **Supervised Learning**

Supervised learning algorithms and supervised learning models make predictions based on labeled training data. Each training sample includes an input and a desired output. A supervised learning algorithm analyzes this sample data and makes an inference – basically, an educated guess when determining the labels for unseen data.

This is the most common and popular approach to machine learning. It's "supervised" because these models need to be fed manually tagged sample data to learn from. Data is labeled to tell the machine what patterns (similar words and images, data categories, etc.) it should be looking for and recognize connections with.

For example, if you want to automatically detect spam, you would need to feed a machine learning algorithm examples of emails that you want classified as spam and others that are important, and should not be considered spam.

The two types of supervised learning tasks: classification and regression.

1.1 Classification in supervised machine learning:

There are a number of classification algorithms used in supervised learning, with Support Vector Machines (SVM) and Naive Bayes among the most common.

In classification tasks, the output value is a category with a finite number of options. For example, with this free pre-trained sentiment analysis model, you can automatically classify data as positive, negative, or neutral.

1.2 Regression in supervised machine learning:

In regression tasks, the expected result is a continuous number. This model is used to predict quantities, such as the probability an event will happen, meaning the output may have any number value within a certain range. Predicting the value of a property in a specific neighborhood or the spread of COVID19 in a particular region are examples of regression problems.

2. Unsupervised Learning

Unsupervised learning algorithms uncover insights and relationships in unlabeled data. In this case, models are fed input data but the desired outcomes are unknown, so they have to make inferences based on circumstantial evidence, without any guidance or training. The models are not trained with the “right answer,” so they must find patterns on their own.

One of the most common types of unsupervised learning is clustering, which consists of grouping similar data. This method is mostly used for exploratory analysis and can help you detect hidden patterns or trends.

For example, the marketing team of an e-commerce company could use clustering to improve customer segmentation. Given a set of income and spending data, a machine learning model can identify groups of customers with similar behaviors.

Segmentation allows marketers to tailor strategies for each key market. They might offer promotions and discounts for low-income customers that are high spenders on the site, as a way to reward loyalty and improve retention.

3. Semi-Supervised Learning

In semi-supervised learning, training data is split into two. A small amount of labeled data and a larger set of unlabeled data.

In this case, the model uses labeled data as an input to make inferences about the unlabeled data, providing more accurate results than regular supervised-learning models.

This approach is gaining popularity, especially for tasks involving large datasets such as image classification. Semi-supervised learning doesn’t require a large number of labeled data, so it’s faster to set up, more cost-effective than supervised learning methods, and ideal for businesses that receive huge amounts of data.

4. Reinforcement Learning

Reinforcement learning (RL) is concerned with how a software agent (or computer program) ought to act in a situation to maximize the reward. In short, reinforced machine learning models attempt to determine the best possible path they should take in a given situation. They do this through trial and error. Since there is no training data, machines learn from their own mistakes and choose the actions that lead to the best solution or maximum reward.

This machine learning method is mostly used in robotics and gaming. Video games demonstrate a clear relationship between actions and results, and can measure success by keeping score. Therefore, they’re a great way to improve reinforcement learning algorithms.

- In this project we are trying to reduce the language barriers among people with a communication technique from amongst speech-trained systems that achieves better performance than those trained with normal speech. Voice emotion recognition is also used in call center applications and mobile wireless communications.

- Voice Classification means categorizing certain sounds/audios in some categories, like environmental sound classification and speech recognition. The task we perform same as in Image classification of cat and dog, Text classification of spam and ham. It is the same applied in voice classification. The only difference is the type of data where we have images, text, and now we have a certain type of voices file of a certain length.

Emotions available:

There are 9 emotions available: "surprise", "angry", "calm" "sad", "disgust", "sad", "fear", "neutral" and “happy”.

Problem Statement

Voice classification using ML

Built a python application that analyses the sentiment behind the tone of the voice and predicts the sentiment involved, you can use an open source dataset for the same.

This project addresses the problem of enhancing voice classification through the application of machine learning (ML) techniques. The objective is to develop a model that can effectively classify varied voice samples, accommodating different speech patterns and environmental conditions. By tackling this issue, the project aims to contribute to the improvement of voice-related applications, such as speech recognition systems, by providing more reliable and adaptable voice classification capabilities

Voice classification and speech recognition which is gaining more popularity and need for it increases enormously. This project attempts to use machine learning deep learning techniques to recognize the emotions from data.

Audio classification employs in industries across different domains like voice lock features, music genre identification, Natural Language classification, Environment sound classification, and to capture and identify different types of sound. It is used in chatbots to provide chatbots with the next level of power.

The following will demonstrate how to apply Machine Learning and Deep Learning techniques to the classification of environmental sounds, specifically focusing on the identification of particular voice or speech.

When given an audio sample in a computer readable format (such as a .wav file) of a few seconds duration, we want to be able to determine if it contains one of the target dataset sounds with a corresponding Classification Accuracy score.

CHAPTER 4

SYSTEM ANALYSIS

4. SYSTEM ANALYSIS

1. Existing System

- The existing has less number of emotions.
- The accuracy is not optimum as expected.
- Existing system has less dataset which leads to less accurate classifications.
- The training technique is not repeated for many time to achieve maximum accuracy

2. Proposed System

- The proposed system is built to achieve maximum accuracy.
- The dataset used in this system is large which the model will train itself to classify the voice again and again.
- This Proposed system has 4 datasets with different voice tones which is large in size.
- This system uses RNN and many feature exactions in details to get each parameters of voice classification

3. Objective of the System

- The objective of system is to achieve maximum accuracy based on dataset.
- To predict accurate result as much as possible to classify the voice or speech.
- To make the clear dataset with the help of Data Augmentation to make our model invariant to those perturbations and enhance its ability to generalize
- The detailed extraction of all the features like Zero Crossing Rate, MFCC, Chroma Vector, RMS(root mean square) value, MelSpectogram to train our model.
- To build and train the best module by fitting all the dataset by various techniques.

CHAPTER 5

REQUIREMENT ANALYSIS

5. REQUIREMENT ANALYSIS

Hardware Requirement Specification

- Voice Recorder.
- Intel i5 or equivalent AMD processor.
- 8GB RAM minimum.
- 256 GB solid state (SSD)hard drive .
- 2 GB GPU.

Software Requirement Specification

- Streamlit.
- GitHub.
- Python 3.10.7.
- Jupyter 6.4.12.
- Required Python linraries like- Numpy,Pandas,Librosa,Keras.
- Dataset with enough size.

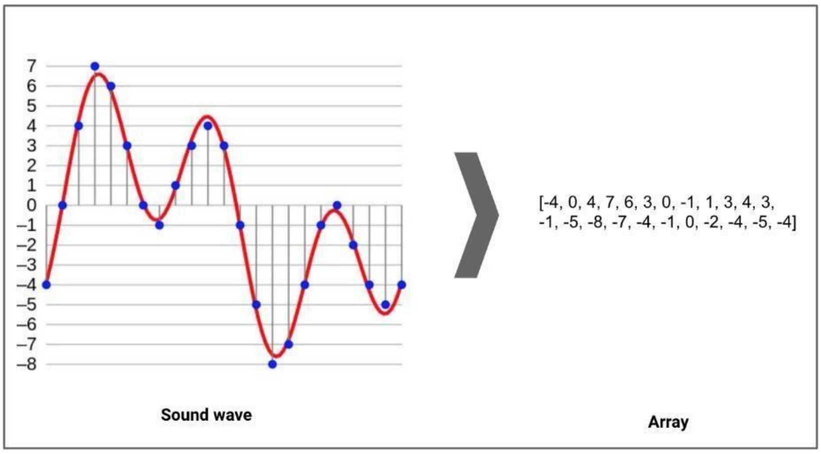
CHAPTER 6

DESIGN ANALYSIS

6. DESIGN & ANALYSIS

The system will demonstrate how to apply Machine Learning and Deep Learning techniques to the classification of environmental sounds, specifically focusing on the identification of particular voice or speech.

When given an audio sample in a computer readable format (such as a .wav file) of a few seconds duration, we want to be able to determine if it contains one of the target dataset sounds with a corresponding Classification Accuracy score**.**

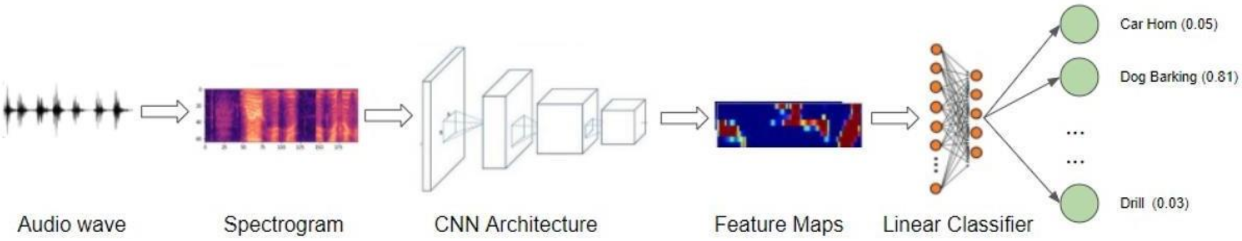


A sound wave, in red, represented digitally, in blue (after sampling and 4-bit quantisation), with the resulting array shown on the right. Original © Aquegg | Wikimedia Commons

Audio/Voice Classification

Just like classifying hand-written digits using the MNIST dataset is considered a 'Hello World"- type problem for Computer Vision, we can think of this application as the introductory problem for audio deep learning.

We will start with sound files, convert them into spectrograms, input them into a CNN plus Linear Classifier model, and produce predictions about the class to which the sound belongs.



Libraries

- **pandas** - Fast, powerful, flexible and easy to use open source data analysis and manipulation library.
- **numpy** - The fundamental package for array computing with Python.
- **os** - The OS module in Python provides functions for interacting with the operating system. The os and os.path modules include many functions to interact with the file system.

- **sys** - The sys module in Python provides various functions and variables that are used to manipulate different parts of the Python runtime environment
- **librosa** - A python package for music and audio analysis. It provides the building blocks necessary to create music information retrieval systems.
- **keras** – Keras is an open-source software library that provides a Python interface for artificial neural networks. It acts as an interface for the TensorFlow library.

Data Preparation

As we are working with four different datasets, so we will be creating a dataframe storing all emotions of the data in dataframe with their paths.

We will use this dataframe to extract features for our model training.

Ravdess Dataframe

Here is the filename identifiers as per the official RAVDESS website:

- Modality (01 = full-AV, 02 = video-only, 03 = audio-only).
- Vocal channel (01 = speech, 02 = song).
- Emotion (01 = neutral, 02 = calm, 03 = happy, 04 = sad, 05 = angry, 06 = fearful, 07 = disgust, 08 = surprised).
- Emotional intensity (01 = normal, 02 = strong). NOTE: There is no strong intensity for the 'neutral' emotion.
- Statement (01 = "Kids are talking by the door", 02 = "Dogs are sitting by the door").
- Repetition (01 = 1st repetition, 02 = 2nd repetition).
- Actor (01 to 24. Odd numbered actors are male, even numbered actors are female).

So, here's an example of an audio filename. 02-01-06-01-02-01-12.mp4 This means the meta data for the audio file is:

- Video-only (02)
- Speech (01)
- Fearful (06)
- Normal intensity (01)
- Statement "dogs" (02)
- 1st Repetition (01)
- 12th Actor (12) - Female (as the actor ID number is even)

TESS dataset

There are a set of 200 target words were spoken in the carrier phrase "Say the word _' by two actresses (aged 26 and 64 years) and recordings were made of the set portraying each of seven emotions (anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral). There are 2800 data points (audio files) in total.

The dataset is organized such that each of the two female actor and their emotions are contain within its own folder. And within that, all 200 target words audio file can be found. The format of the audio file is a WAV format.

Crema-D Dataset

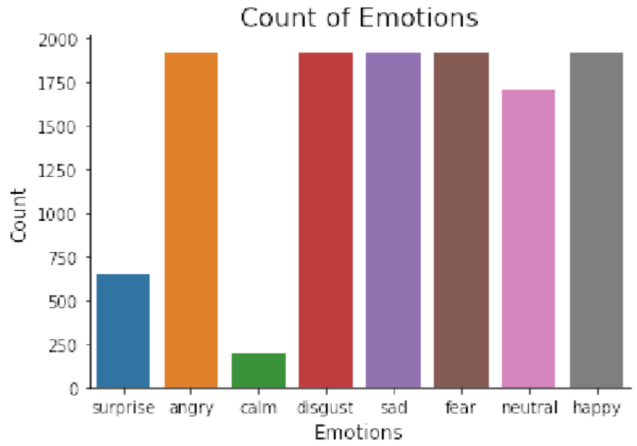
CREMA-D is an emotional multimodal actor data set of 7,442 original clips from 91 actors. These clips were from 48 male and 43 female actors between the ages of 20 and 74 coming from a
variety of races and ethnicities (African America, Asian, Caucasian, Hispanic, and Unspecified).

Actors spoke from a selection of 12 sentences. The sentences were presented using one of six different emotions (Anger, Disgust, Fear, Happy, Neutral, and Sad) and four different emotion levels (Low, Medium, High, and Unspecified).

Participants rated the emotion and emotion levels based on the combined audio visual presentation, the video alone, and the audio alone. Due to the large number of ratings needed, this effort was crowd-sourced and a total of 2443 participants each rated 90 unique clips, 30 audio, 30 visual, and 30 audio-visual. 95% of the clips have more than 7 ratings.

Data Visualization and Exploration

Plotting the count of each emotions of dataset.



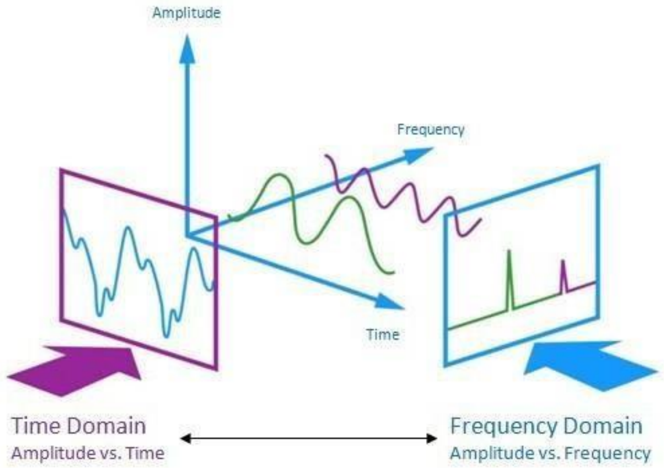
We can also plot waveplots and spectrograms for audio signals:

Waveplots - Waveplots let us know the loudness of the audio at a given time.
Spectrograms - A spectrogram is a visual representation of the spectrum of frequencies of sound or other signals as they vary with time. It’s a representation of frequencies changing with respect to time for given audio/music signals.

Feature Extraction

Extraction of features is a very important part in analyzing and finding relations between different things. As we already know that the data provided of audio cannot be understood by the models directly so we need to convert them into an understandable format for which feature extraction is used.

The audio signal is a three-dimensional signal in which three axes represent time, amplitude and frequency.



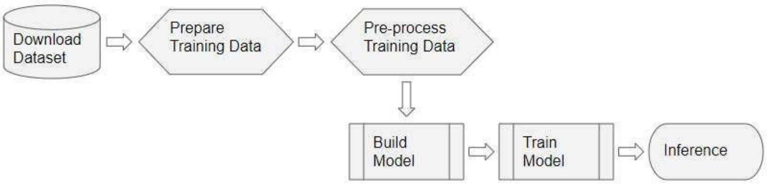
In this project i am not going deep in feature selection process to check which features are good for our dataset rather i am only extracting 5 features:

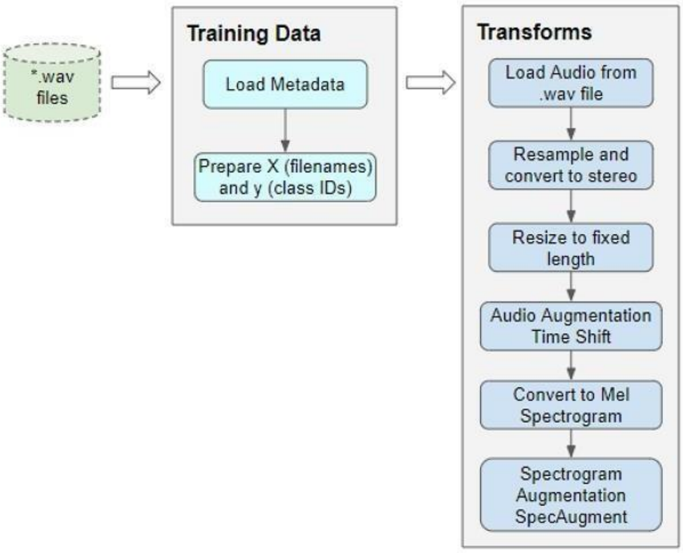
- Zero Crossing Rate
- Chroma_stft
- MFCC
- RMS(root mean square) value
- MelSpectogram to train our model.

Next the Data preparation and Modelling is a essential part of the system now we need to normalize and split our data for training and testing.

Prepare training data :

As for most deep learning problems, we will follow these steps:




*.wav files
Training Data
Load Metadata
Prepare X (filenames) and y (class IDs)
Transforms
Load Audio from .wav file
Resample and convert to stereo
Resize to fixed length
Audio Augmentation Time Shift
Convert to Mel Spectrogram
Spectrogram Augmentation SpecAugment

Pre-processing the training data for input to our model

CHAPTER 7

IMPLEMENTATION

7. IMPLEMENTATION

Implementation is the stage where the theoretical design is turned into a working system. The most crucial stage in achieving a new successful system and in giving confidence on the new system for the users that it will work efficiently and effectively.

The system can be implemented only after thorough testing is done and if it is found to work according to the specification. It involves careful planning, investigation of the current system and it constraints on implementation, design of methods to achieve the change over and an evaluation of change over methods a part from planning.

Two major tasks of preparing the implementation are education and training of the users and testing of the system. The more complex the system being implemented, the more involved will be the system analysis and design effort required just for implementation.

The implementation phase comprises of several activities. The required hardware and software acquisition is carried out. The system may require some software to be developed. For this, programs are written and tested. The user then changes over to his new fully tested system and the old system is discontinued.

TESTING

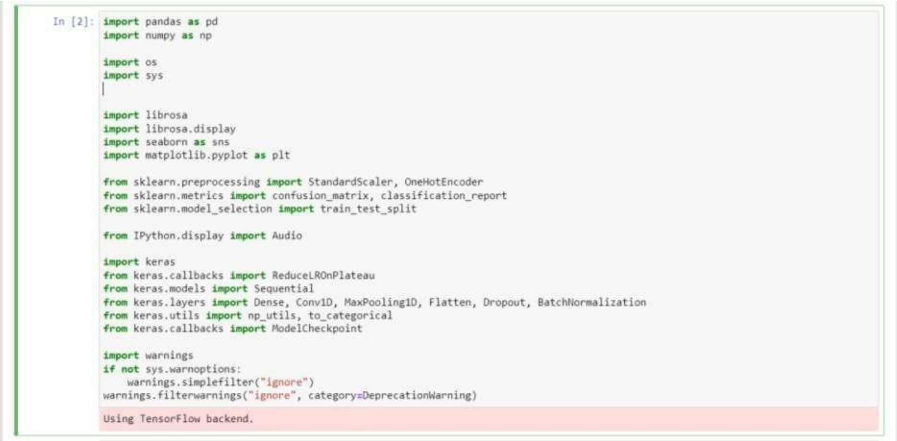
The testing phase is an important part of software development. It is the Information zed system will help in automate process of finding errors and missing operations and also a complete verification to determine whether the objectives are met and the user requirements are satisfied. Software testing is carried out in three steps:

1. The first includes unit testing, where in each module is tested to provide its correctness, validity and also determine any missing operations and to verify whether theobjectives have been met. Errors are noted down and corrected immediately.

2. Unit testing is the important and major part of the project. So errors are rectified easily in particular module and program clarity is increased. In this project entire system is dividedinto several modules and is developed individually. So unit testing is conducted to individual modules.

3. The second step includes Integration testing. It need not be the case, the software whose modules when run individually and showing perfect results, will also show perfect results when run as a whole.

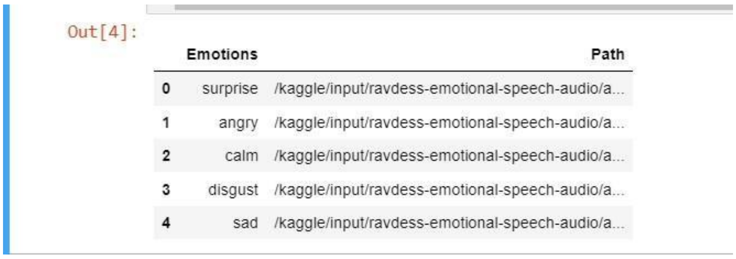
Importing Libraries



```
In [2]: import pandas as pd
        import numpy as np

        import os
        import sys

        import librosa
        import librosa.display
        import seaborn as sns
        import matplotlib.pyplot as plt

        from sklearn.preprocessing import StandardScaler, OneHotEncoder
        from sklearn.metrics import confusion_matrix, classification_report
        from sklearn.model_selection import train_test_split

        from IPython.display import Audio

        import keras
        from keras.callbacks import ReduceLROnPlateau
        from keras.models import Sequential
        from keras.layers import Dense, Conv1D, MaxPooling1D, Flatten, Dropout, BatchNormalization
        from keras.utils import np_utils, to_categorical
        from keras.callbacks import ModelCheckpoint

        import warnings
        if not sys.warnoptions:
            warnings.simplefilter("ignore")
        warnings.filterwarnings("ignore", category=DeprecationWarning)

        Using TensorFlow backend.
```

Paths for data

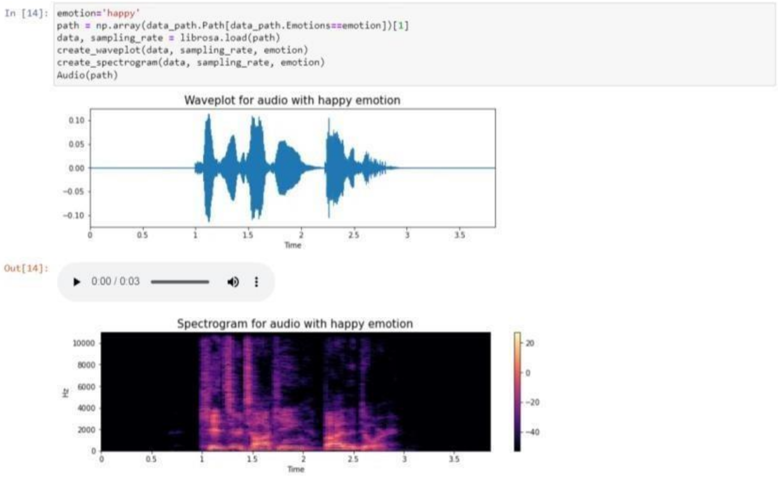
Ravdess = "/kaggle/input/ravdess-emotional-speech-audio/audio_speech_actors_01-24/"
Crema = "/kaggle/input/cremad/AudioWAV/"
Tess = "/kaggle/input/toronto-emotional-speech-set-tess/tess toronto emotional speech set data/TESS Toronto emotional speech set data/"
Savee = "/kaggle/input/surrey-audiovisual-expressed-emotion-savee/ALL/"



Out[4]:

	Emotions	Path
0	surprise	/kaggle/input/ravdess-emotional-speech-audio/a...
1	angry	/kaggle/input/ravdess-emotional-speech-audio/a...
2	calm	/kaggle/input/ravdess-emotional-speech-audio/a...
3	disgust	/kaggle/input/ravdess-emotional-speech-audio/a...
4	sad	/kaggle/input/ravdess-emotional-speech-audio/a...

Plotting waveplots and spectrograms for audio signals



```
In [14]: emotion='happy'
         path = np.array(data_path.Path[data_path.Emotions==emotion])[1]
         data, sampling_rate = librosa.load(path)
         create_waveplot(data, sampling_rate, emotion)
         create_spectrogram(data, sampling_rate, emotion)
         Audio(path)
```

Data Augmentation

- Data augmentation is the process by which we create new synthetic data samples by adding small perturbations on our initial training set.
- To generate syntactic data for audio, we can apply noise injection, shifting time, changing pitch and speed.
- The objective is to make our model invariant to those perturbations and enhance its ability to generalize.
- In order to this to work adding the perturbations must conserve the same label as the original training sample.
- In images data augmentation can be performed by shifting the image, zooming, rotating ...

Feature Extraction

Extraction of features is a very important part in analyzing and finding relations between different things. As we already know that the data provided of audio cannot be understood by the models directly so we need to convert them into an understandable format for which feature extraction is used.

The audio signal is a three-dimensional signal in which three axes represent time, amplitude and frequency.

As stated there with the help of the sample rate and the sample data, one can perform several transformations on it to extract valuable features out of it.

Zero Crossing Rate : The rate of sign-changes of the signal during the duration of a particular frame.

Energy : The sum of squares of the signal values, normalized by the respective frame length.

Entropy of Energy : The entropy of sub-frames' normalized energies. It can be interpreted as a measure of abrupt changes.

Spectral Centroid : The center of gravity of the spectrum.

Spectral Spread : The second central moment of the spectrum.

Spectral Entropy : Entropy of the normalized spectral energies for a set of sub-frames.

Spectral Flux : The squared difference between the normalized magnitudes of the spectra of the two successive frames.

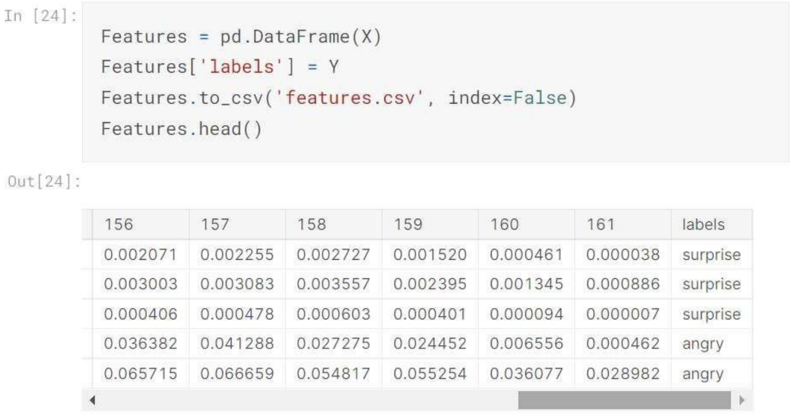
Spectral Rolloff : The frequency below which 90% of the magnitude distribution of the spectrum is concentrated.

MFCCs Mel Frequency Cepstral Coefficients form a cepstral representation where the frequency bands are not linear but distributed according to the mel-scale.

Chroma Vector : A 12-element representation of the spectral energy where the bins represent the 12 equal-tempered pitch classes of western-type music (semitone spacing).

Chroma Deviation : The standard deviation of the 12 chroma coefficients.

In final stage the Data preparation and Modelling is a essential part of the system now we need to normalize and split our data for training and testing.



In [24]:

```
Features = pd.DataFrame(X)
Features['labels'] = Y
Features.to_csv('features.csv', index=False)
Features.head()
```

Out[24]:

156	157	158	159	160	161	labels
0.002071	0.002255	0.002727	0.001520	0.000461	0.000038	surprise
0.003003	0.003083	0.003557	0.002395	0.001345	0.000886	surprise
0.000406	0.000478	0.000603	0.000401	0.000094	0.000007	surprise
0.036382	0.041288	0.027275	0.024452	0.006556	0.000462	angry
0.065715	0.066659	0.054817	0.055254	0.036077	0.028982	angry

The same can be implemented by converting into H5 model and creating a frontend in streamlit

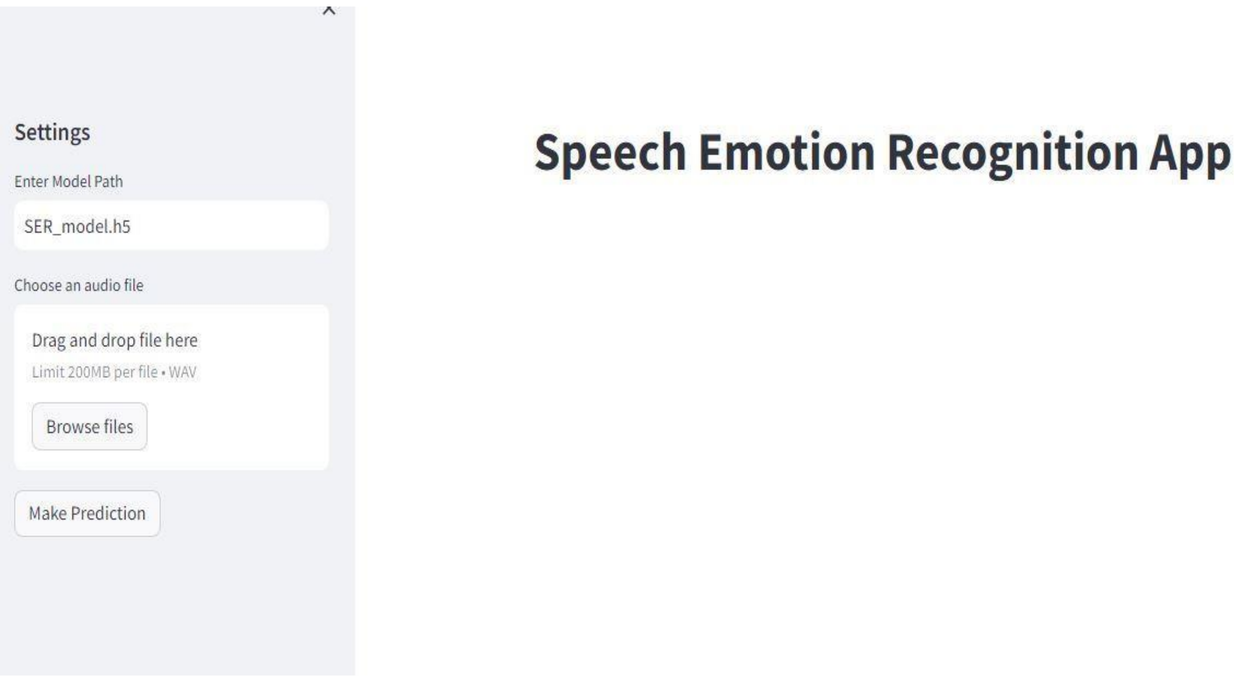
Later the backend can be hosted by linking Github profile with Streamlit

In this way the virtual workspace can also be used and the snapshot of so is used in the further

CHAPTER 8

SNAPSHOTS

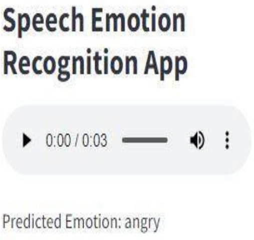
8. SNAPSHOTS



Speech Emotion Recognition App

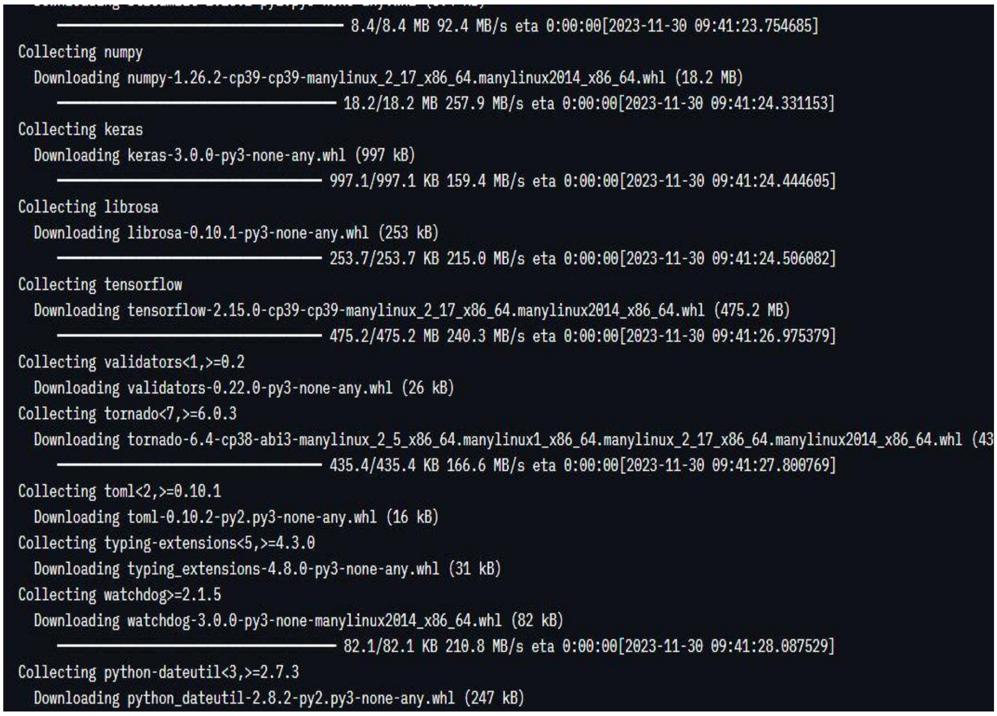
Enter Model Path: SER_model.h5

Choose an audio file

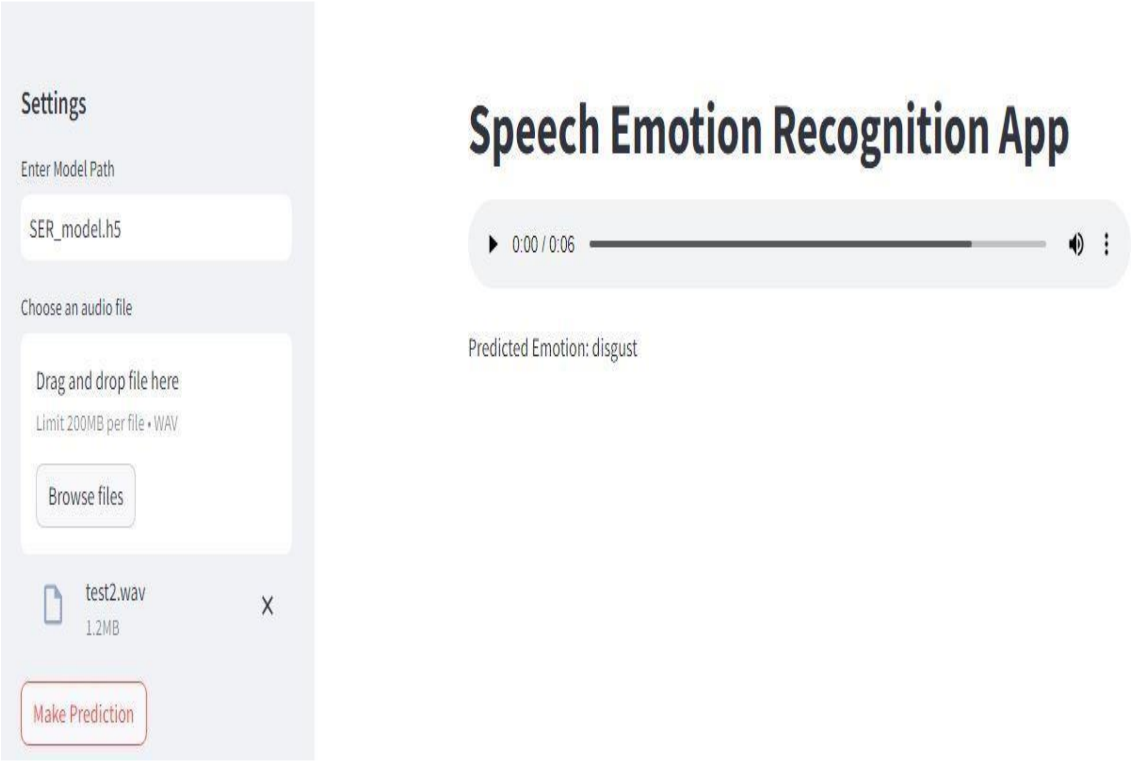


Speech Emotion Recognition App

Predicted Emotion: angry



```
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 8.4/8.4 MB 92.4 MB/s eta 0:00:00[2023-11-30 09:41:23.754685]
Collecting numpy
  Downloading numpy-1.26.2-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (18.2 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 18.2/18.2 MB 257.9 MB/s eta 0:00:00[2023-11-30 09:41:24.331153]
Collecting keras
  Downloading keras-3.0.0-py3-none-any.whl (997 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 997.1/997.1 KB 159.4 MB/s eta 0:00:00[2023-11-30 09:41:24.444605]
Collecting librosa
  Downloading librosa-0.10.1-py3-none-any.whl (253 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 253.7/253.7 KB 215.0 MB/s eta 0:00:00[2023-11-30 09:41:24.506082]
Collecting tensorflow
  Downloading tensorflow-2.15.0-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (475.2 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 475.2/475.2 MB 240.3 MB/s eta 0:00:00[2023-11-30 09:41:26.975379]
Collecting validators<1,>=0.2
  Downloading validators-0.22.0-py3-none-any.whl (26 kB)
Collecting tornado<7,>=6.0.3
  Downloading tornado-6.4-cp38-abi3-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux2014_x86_64.whl (43
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 435.4/435.4 KB 166.6 MB/s eta 0:00:00[2023-11-30 09:41:27.800769]
Collecting toml<2,>=0.10.1
  Downloading toml-0.10.2-py2.py3-none-any.whl (16 kB)
Collecting typing-extensions<5,>=4.3.0
  Downloading typing_extensions-4.8.0-py3-none-any.whl (31 kB)
Collecting watchdog>=2.1.5
  Downloading watchdog-3.0.0-py3-none-manylinux2014_x86_64.whl (82 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 82.1/82.1 KB 210.8 MB/s eta 0:00:00[2023-11-30 09:41:28.087529]
Collecting python-dateutil<3,>=2.7.3
  Downloading python_dateutil-2.8.2-py2.py3-none-any.whl (247 kB)
```



Settings

Enter Model Path

SER_model.h5

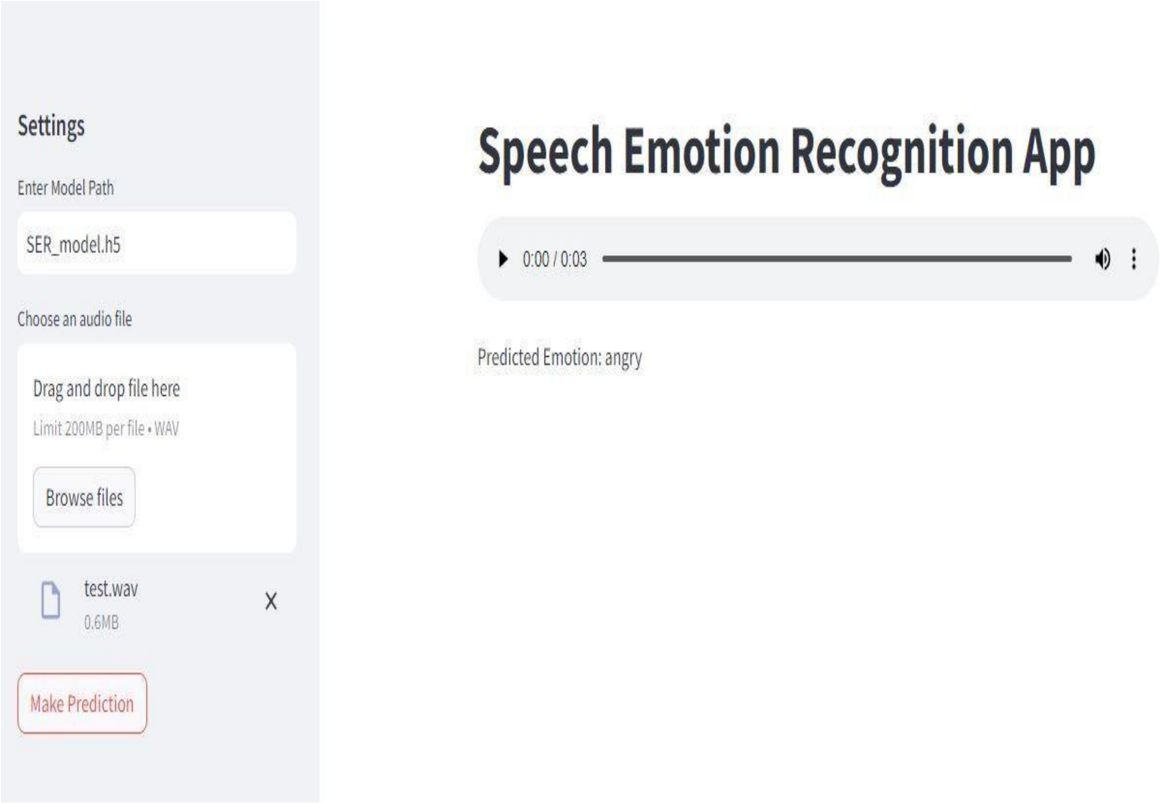
Choose an audio file

test2.wav

1.2MB

Speech Emotion Recognition App

Predicted Emotion: disgust



Settings

Enter Model Path

SER_model.h5

Choose an audio file

test.wav

0.6MB

Speech Emotion Recognition App

Predicted Emotion: angry

CHAPTER 9

CONCLUTION

9. CONCLUTION

The package was designed in such a way that future modifications can be done easily. The following conclusions can be deduced from the development of the project:

- Automation of the entire system improves the efficiency
- It provides a friendly graphical user interface which proves to be better when compared to the existing system.
- It gives appropriate access to the authorized users depending on their permissions.
- It effectively overcomes the delay in communications.
- Updating of information becomes so easier
- System security, data security and reliability are the striking features.
- The System has adequate scope for modification in future if it is necessary.
- Our model is more accurate in predicting surprise, angry emotions and it makes sense also because audio files of these emotions differ to other audio files in a lot of ways like pitch, speed etc..
- We overall achieved 61% accuracy on our test data and its decent but we can improve it more by applying more augmentation techniques and using other feature extraction methods.
- We can visualize any audio in the form of a waveform.
- MFCC method is used to extract important features from audio files.
- Scaling the audio samples to a common scale is important before feeding data to the model to understand it better.
- We can build a RNN model to classify audios.

10. REFERENCE

- An Audio Classification Approach Based on Machine

Learning Date of Conference: 12-13 January 2019
Date Added to IEEE *Xplore*: 21
March 2019 INSPEC Accession
Number: 18529924 DOI:
10.1109/ICITBS.2019.00156
Publisher: IEEE

- Support Vector Machine based Voice Activity Detection

Published in: 2006 International Symposium on Intelligent Signal Processing and
Communications Date of Conference: 12 December 2005 - 15 December 2006
Date Added to IEEE *Xplore*: 29
May 2007 INSPEC Accession
Number: 9505927 DOI:
10.1109/ISPACS.2006.364896
Publisher: IEEE

- Sound Classification using Deep Learning

https://mikesmales.medium.com/sound-classification-using-deep-learning-8bc2aa1990b7

- Audio Deep Learning Made Simple: Sound Classification, Step-by-Step, Published in Towards Data Science

https://towardsdatascience.com/audio-deep-learning-made-simple-sound-classification-step-by- step-cebc936bbe5

- Machine Learning for Audio

Classification Author : Willies Ogola
https://www.section.io/engineering-education/machine-learning-for-audio-classification/

- Implementing Audio Classification Project Using Deep Learning

Raghav Agrawal — Published On March 16, 2022 and Last Modified On April 7th, 2022
https://www.analyticsvidhya.com/blog/2022/03/implementing-audio-classification-project-using- deep-learning/

- Usage of streamlite-
  https://www.latentview.com/data-engineering-lp/introduction-to-streamlit/

- Hosting Streamlite using Github-
  https://www.youtube.com/watch?v=2VB09IT3r44