SPEECH EMOTION BASED RECOGNITION SYSTEM

A Major Project Report submitted in partial fulfilment of the Bachelor of Technology degree requirement.in

Computer Science and Engineering

By

SALONI SINGH (181246)

UNDER THE SUPERVISION OF

DR. SURJEET SINGH

Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology, Waknaghat, 173234, Himachal Pradesh, IND

Certificate

Candidate's Declaration

I hereby declare that the work presented in this report entitled "SPEECH EMOTION BASED RECGOGNITION SYSTEM" in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering/Information Technology submitted in the department of ComputerScience &amp; Engineering and Information Technology, Jaypee University ofInformation Technology Waknaghat is an authentic record of my own work carried out over a period from August 2021 to December 2021 under the supervision of Dr. Surjeet Singh Assistant Professor(SG) Department of Computer Science and Engineering.

The matter embodied in the report has not been submitted for the award of any other degree or diploma.

SALONI SINGH
Saloni Singh (181246)

This is to certify that the above statement made by the candidate is true to the bestof my knowledge.

(SupervisorSignature)
Dr. Surjeet Singh
Assistant
Professor(SG)
Computer Science
andEngineering
Dated:

AKCNOWLEDGEMENT

To begin, I want to express my heartfelt gratitude to almighty God for His heavenly grace, which has enabled us to successfully complete the project work.

Supervisor Dr. SURJEET SINGH, Assistant Professor (SG), Department of CSE Jaypee University of Information Technology, Waknaghat, deserves my deepest gratitude. His never-ending patience, intellectual direction, constant encouragement, constant and energetic supervision, constructive criticism, helpful suggestions, and reading numerous poor versions and revising them at every level allowed this project to be completed.

I'd like to thank Dr. SURJEET SINGH, Department of CSE, for his invaluable assistance in completing my project.I'd also like to express my gratitude to everyone who has assisted me in makingthis project a success, whether directly or indirectly. In this unique situation, I'dwant to express my gratitude to the different staff members, both teaching and non-teaching, who have provided me with valuable assistance and assisted my project. Finally, I must express my gratitude for my parents' unwavering support andpatience.

Saloni Singh

ABSTRACT

The emotions from signals of speech have been recognized by the tricky module in CAS (Computer Aided Services). For extraction of emotions from the signals, we will be using several schemes, comprising of classification and speech analysis methods.

The project proposes a model in which the algorithm of Machine Learning is used to detect the various emotions of the individual on the basis of his pitch, tone and frequency of voice. This manuscript summaries methodologies & analyses some recent literature in which prevalent models for emotion recognition based on speech have been applied. Talking about the system it will consist of voice activity detection, speech segmentation, feature extraction and the emotion frequency to be statistically analyzed.

Four categories of experiments were conducted using pre-recorded datasets as well as real-time recording. The five different emotions are namely happy, disgust, sad, fear and anger.

Table of Content

CHAPTER 1

1. Introduction

Speech Emotion Recognition (SER) is the process of attempting to recognize people's emotions as well as the circumstances around their speech. This is because the truth often reflects the basic feelings of tone and tone of voice. It is the situation that animals such as horses and dogs use for understanding human emotions.

The recognition is difficult as the emotions are subdued and the sound of the annotations is challenging.In this project, we have used librosa libraries, soundfile, and sklearn libraries (among others) to create a model using MLP Classifier. Detection of emotions from audio files is possible due to this. We will upload the data, extract the it’s features, and then splitting of database into sets of testing and training. After that, we will launch the MLP Classifier and it’s training is done. Finally, the model’s accuracy is calculated.

We’ll use the RAVDESS dataset for this research; i.e., the Ryerson Audio-Visual Database of Emotional Speech and Song dataset, and it's available for free to download. There are 2880 files in the database, all of which have been graded on emotional, dynamic, and realistic performance. The entire database is 1.09GB from 24 characters, but we have reduced the sample size to all files.

Artificial Intelligence’s (AI’s) growth has been accelerated from several years now. AI, once topic studied only by engineers and scientists, has now reached the home of the each and every person as an intelligent program. The development of AI has created many technologies that include Human Computing (HCI). Because HCI is the ultimate AI available to millions of people, aiming to enhance and improve HCI approaches is crucial. Touch communication, movement, hand gestures, voice, and facial expressions are some of the available HCI approaches. Intelligent voice-based devices are gaining popularity in a variety of applications among the many methodologies. A computer programmer must thoroughly grasp a person's speech in order to accept exact instructions in a word-based software. Speech Processing is the field of study which comprises of three components:

- Speaker Identification
- Speech Emotion Detection
- Speech Recognition

Detection of Speech emotion (SED) is a challenge to use because of its complexity among various components. In addition, what a computer system(intelligent) requires is the program mimic human behavior. A surprisingly unique feature for people is the ability to turn conversations of speaker and listener based on their emotional state. Speech sensory detection may be constructed as a separation problem which are solved by various algorithms of machine learning. This This project has various methods discussed in detail and tests performed as part of the implementation of the Speech Emotion Detection System.

.

2. Objective of the Major Project

Machine Learning is a buzz word in the modern era of technology. People these days have a misconception regarding machine learning, which includes learning the algorithms. However, in actual it’s not about just algorithms. Algorithms are the beginning step towards Machine Learning. The main motive Machine Learning model is learning patterns that generalize well for unseen data instead of just memorizing the data which was shown during training.

The later part is Optimization, which includes increasing efficiency of your model. Our primary goal is to achieve an effective model for which various iterations and regressive efforts are required and done. Our purpose of using Machine Learning for our model is because of the various advantages it provides, the major advantage being – human like intelligence to a machine. The various stages through which the model development cycle goes, starts from data collection to model building. Our model requires data set from the user, analyzing it and generating an output. For this purpose, the best method would be, none other than, Machine Learning. To measure predictive accuracy many matrices are used in ML and the choice of these matrices depends upon the task along with the reviewing of the matrices for better performance. The choice of accuracy metric is dependent on the Machine Learning task. It’s important to review these metrics to decide if our model is performing well or not.

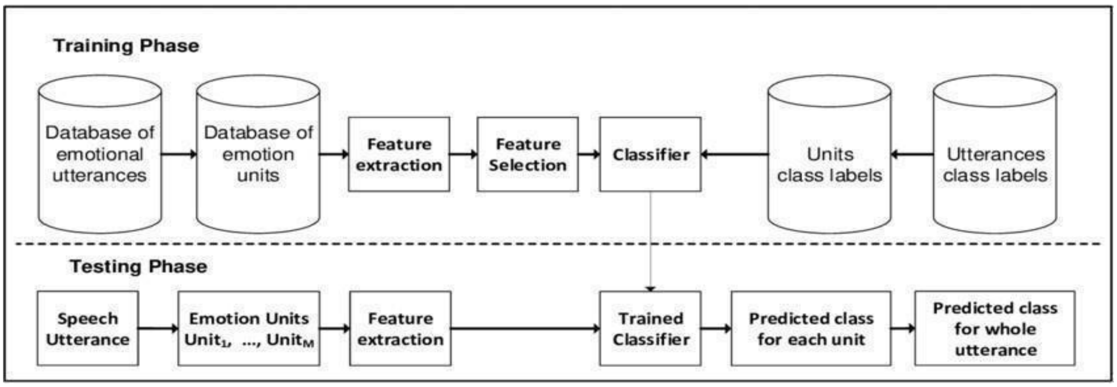
3. Problem Statement

Towards the study of emotional content of speech signals, there has been an increasing attention due to which many proposals related to spoken utterance have come out. The three emotional aspects of design are as defined-

- creation of a classification scheme that is appropriate.
- a database that has been properly prepared.

The proposed system for detecting the emotional state is shown in Figure. This system is composed of two stages: the training phase and the testing phase. In the first stage utteranceis segmented into its emotion units. Then acoustic features extracted from each emotion unit. The third step is implementing a feature selection method to use only the most discriminative acoustic features. It is assumed that each emotion unit has the same emotion class as the utterance they belong to. The final step is to train the proposed classifier to learnthe relationship between acoustic features extracted from emotion unit and the emotional state of this unit. The second stage is the testing phase; this stage is used to predict the emotional state of a new utterance using the trained SER system.

Fig.1 Problem Statement



The special features of the project include:

1.) A user-friendly website which provides a
description of all the 8 emotions to berecognized.

2.) Double masking for more accuracy in the project.

3.) A wholesome process for recognition of the emotion at just one click.

CHAPTER 2

LITERATURE SURVEY

M.Ayadi ,et al [1] suggested that Recently, increasing attention has been directed to the study of emotional content of speech signals, therefore, more systems have been proposed to identify the emotional content of spoken speech. This paper is a survey of the emotional separation of speech that discusses three key elements of the formation of the emotional awareness system. The first is to select the appropriate elements to represent the speech. The second issue is the design of the appropriate partition system and the third issue is the proper setting of the heart-to-heart speech website to test system performance. Conclusions regarding the performance and limitations of current speech recognition systems are discussed in the final section of this survey.

R.Khalil,et al.[2] describes numerous strategies that can be used in the literature of speech emotion recognition to extract emotions from signals, including many well-established speech analysis and classification techniques. Deep Learning approaches have lately been recommended as a replacement for classic SER techniques. Their paper provides an introduction of Deep Learning techniques and covers some current research that uses such approaches to recognise speech-based emotions. The review discusses the databases utilised, the emotions retrieved, contributions to speech emotion identification, and its limitations.

Schuller et al.[3] explains continuous hidden Markov models, we demonstrate voice emotion recognition. The spreading of two approaches is compared. In the first technique, Gaussian mixture models are used to classify a global statistical framework of an utterance using derived features of the raw pitch and energy contour of the speech signal. A second technique adds temporal complexity by employing continuous hidden Markov models with low-level instantaneous features instead of global statistics to include many states. The paper explains the design of functional recognition engines and the outcomes obtained in comparison to the stated alternatives.

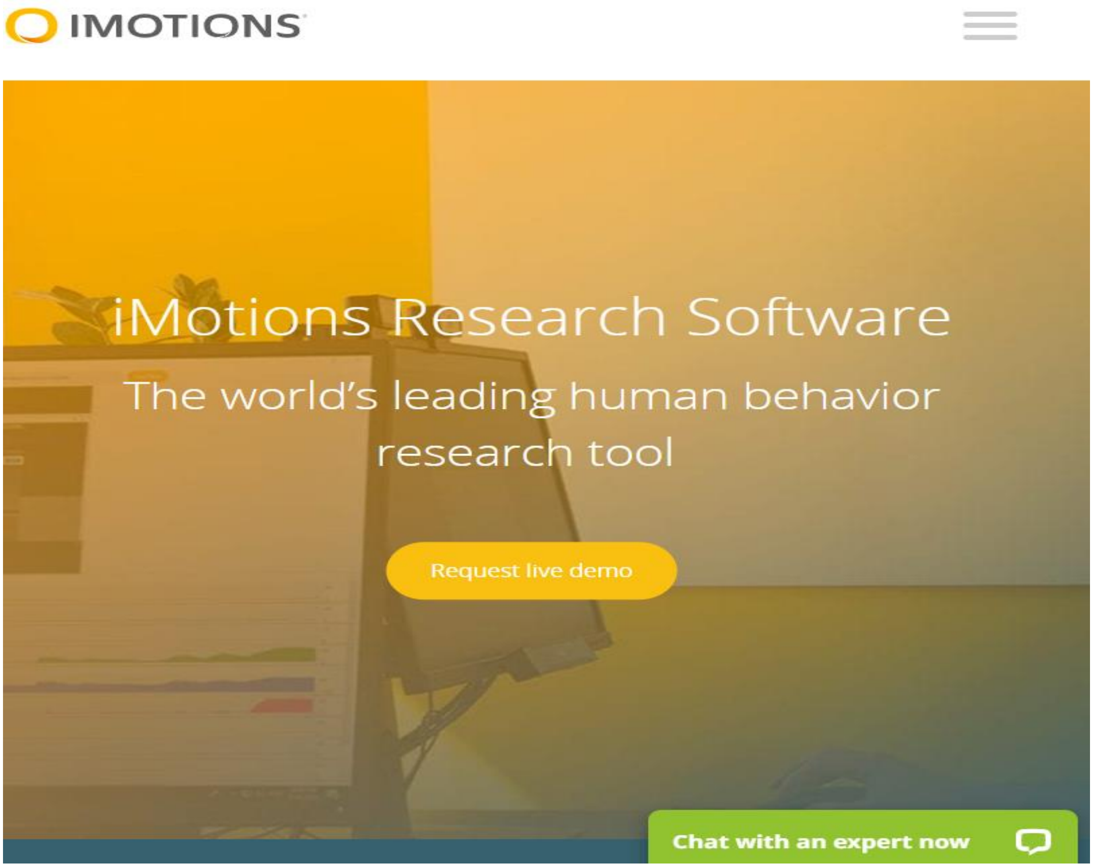
New et al.[4] States Statistics of basic frequency, energy contour, duration of stillness, and voice quality are prominent characteristics used in speech signal emotion identification. When more than two types of emotion need to be categorised, however, the performance of systems using these features falls substantially. This study proposes a text-independent technique for voice emotion categorization. Short time log frequency power coefficients (LFPC) are used to represent speech signals, and a discrete hidden Markov model (HMM) is used as the classifier in the proposed method.

Huang et al.[5] states Convolutional Neural Networks may infer a hierarchical representation of input data, making categorization easier. We propose employing semi-CNN to learn affect-salient features for Speech Emotion Recognition (SER). There really are two steps of semi-CNN training. Unlabeled samples are used in the first stage to train a contractive convolutional neural network with reconstruction penalization to learn candidate features. In the second stage, the candidate features are then fed into a semi-CNN, which uses a new objective function to encourage feature saliency, orthogonality, and discrimination when training affect-salient, discriminative features.

We have researched the above techniques/approaches to implement our project after analyzing existing techniques/approach in the healthcare and technology domain. On the basis of our research, we have created our own new approach taking references from our research part and successfully implemented our project and later we have compared it with the existing technology working on similar kind of projects.

2 . Softwares and Websites

1) Emotient



iMotions Research Software

The world's leading human behavior research tool

Fig.2 Website of Emotient

This website can be used for a variety of reasons it can grasp how people feel for a varety

Of things such as

a certain ad

a certain product

a certain tv show or a movies

This website is also very helpul in healthcare, it is also very efficient for big tech companies

Such as apple to analysis of shopping experience of people buying stuff from apple store.

2) Affectiva

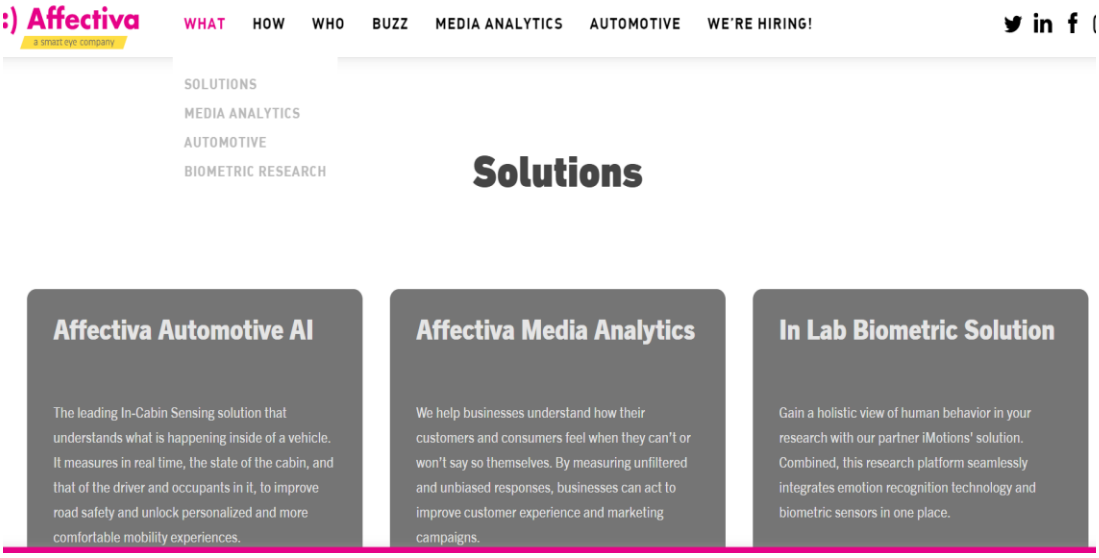
In-Cabin Sensing's leading solution that understands what's going on inside the car. It measures

real-time, cabinet condition, as well as the driver and its occupants, in order to improve road safety and unlock personal experience and comfort of travel. We help businesses understand how their customers and consumers feel when they can or would not say so themselves. By balancing non-critical and impartial responses, businesses can work to improve customer experience and marketing campaigns. Get a complete overview of human behavior in your research with the iMotions Partner solution. Collectively, this field of study integrates seamless biometric sensory and sensory technology into a single area. Affectiva focuses on developing its technology in critical comprehension, automotive, media, audience and customer statistics, community robots. Educational researchers should contact iMotions, which has our technology integrated with their platform.

This website is little different from speech recognition it has analyzed 3,289,274 faces till date.

This website owners provide sdks and apis for mobile devs to understand and document expre

Ssions for later use.



Solutions

Affectiva Automotive AI

The leading In-Cabin Sensing solution that understands what is happening inside of a vehicle. It measures in real time, the state of the cabin, and that of the driver and occupants in it, to improve road safety and unlock personalized and more comfortable mobility experiences.

Affectiva Media Analytics

We help businesses understand how their customers and consumers feel when they can't or won't say so themselves. By measuring unfiltered and unbiased responses, businesses can act to improve customer experience and marketing campaigns.

In Lab Biometric Solution

Gain a holistic view of human behavior in your research with our partner iMotions' solution. Combined, this research platform seamlessly integrates emotion recognition technology and biometric sensors in one place.

Fig.3 Web page of affectiva

Applications:

1 . Road safety

Ai understanding whats happening inside a vehicle. Measuring real time state inside the Vehicle to ensure safety and personalized experience.

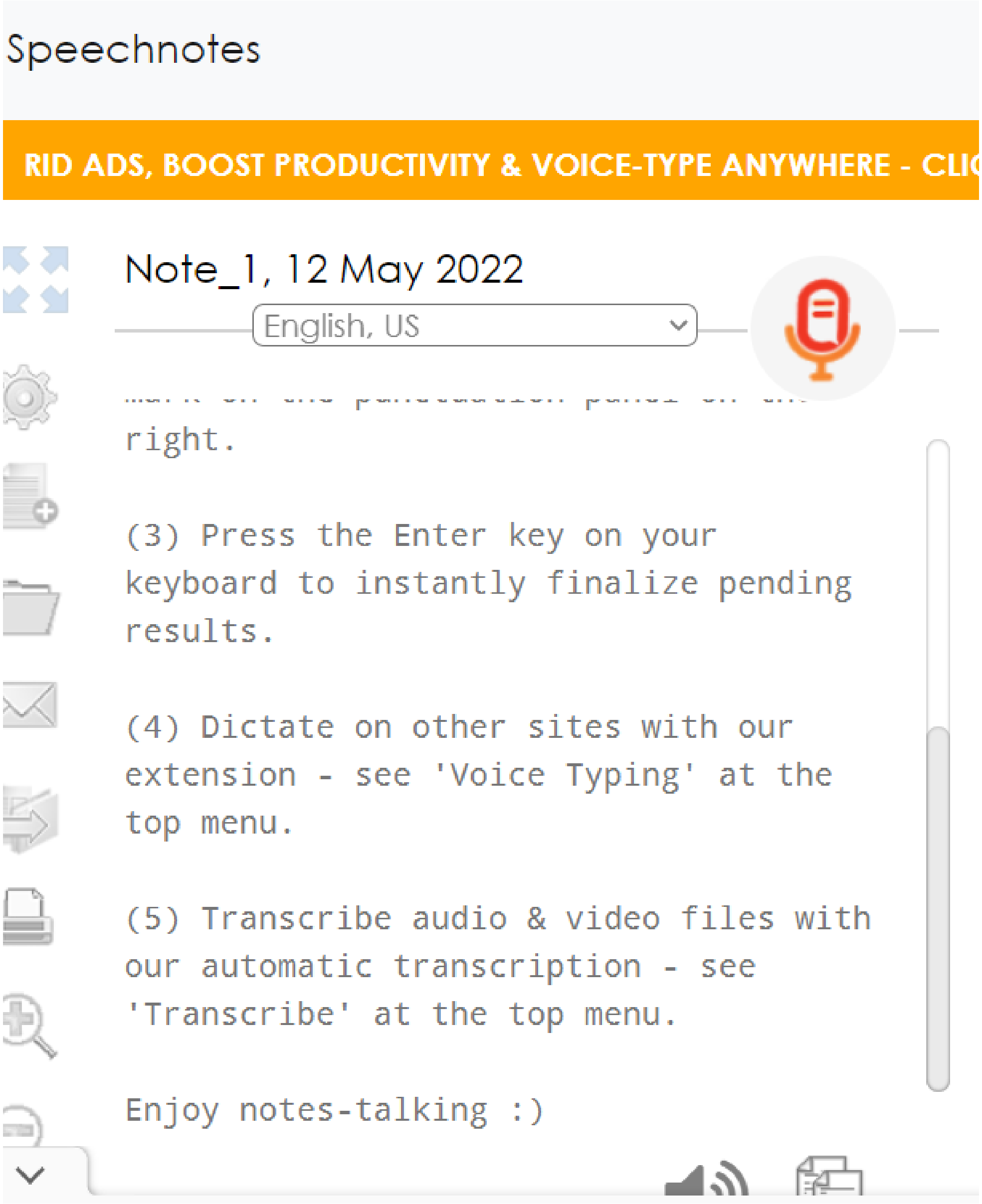
2 . Biometric solutions

Gain a close view of human behavior and implement it in your research.

3 . Media analysis

Helps various companies to understand what they don't personally know about customers By the help of unfiltered reviews they can act upon what is supposed to change or improvised.

3. Speechnotes



Note_1, 12 May 2022

English, US

right.

(3) Press the Enter key on your keyboard to instantly finalize pending results.

(4) Dictate on other sites with our extension - see 'Voice Typing' at the top menu.

(5) Transcribe audio & video files with our automatic transcription - see 'Transcribe' at the top menu.

Enjoy notes-talking :)

Fig.4 Speechnote website outlook

this application website lets you to voice type and it lets you establish a habbit where

You can type and dictate at a very good speed simultaneously.it also helps you to edit

Your text notes through speech only which makes it a very efficient app for students.

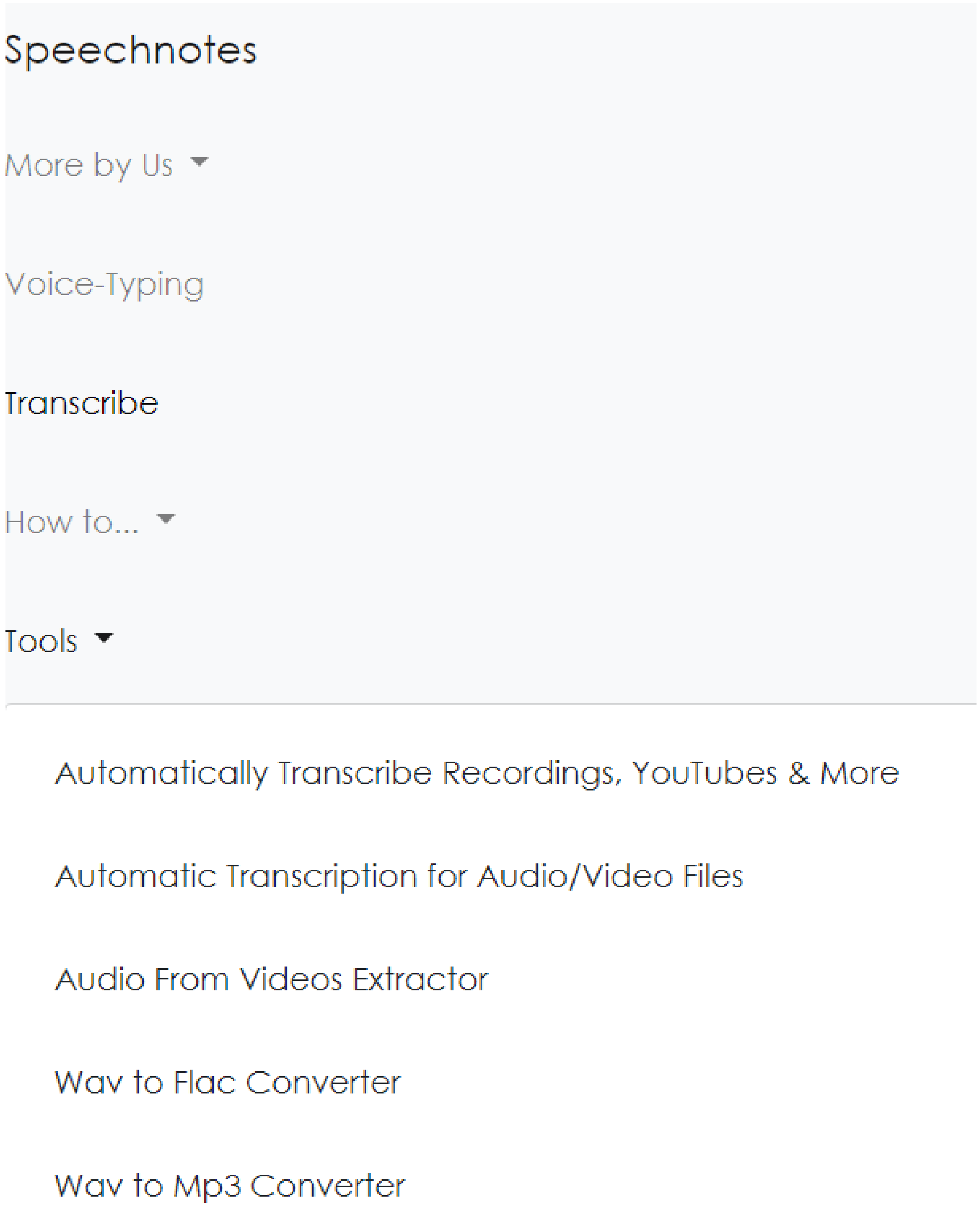

Speechnotes
More by Us
Voice-Typing
Transcribe
How to...
Tools
Automatically Transcribe Recordings, YouTubes & More
Automatic Transcription for Audio/Video Files
Audio From Videos Extractor
Wav to Flac Converter
Wav to Mp3 Converter

Fig.5 Feautres in website

it comes with various features and tools for optimal user experience.

And also comes with a user guide on how to use!

3.Voice recognition softwares

1) Dragon professional.

This website provides a platform for professionals to create reports,forms and through speech.

It provides solutions for various fields:

1) law enforcement
2) heathcare
3) financial services
4) Telecommunication
5) retail
6) government
7) Insurance
8) Social services
9) travel and hospitality
10) utilities

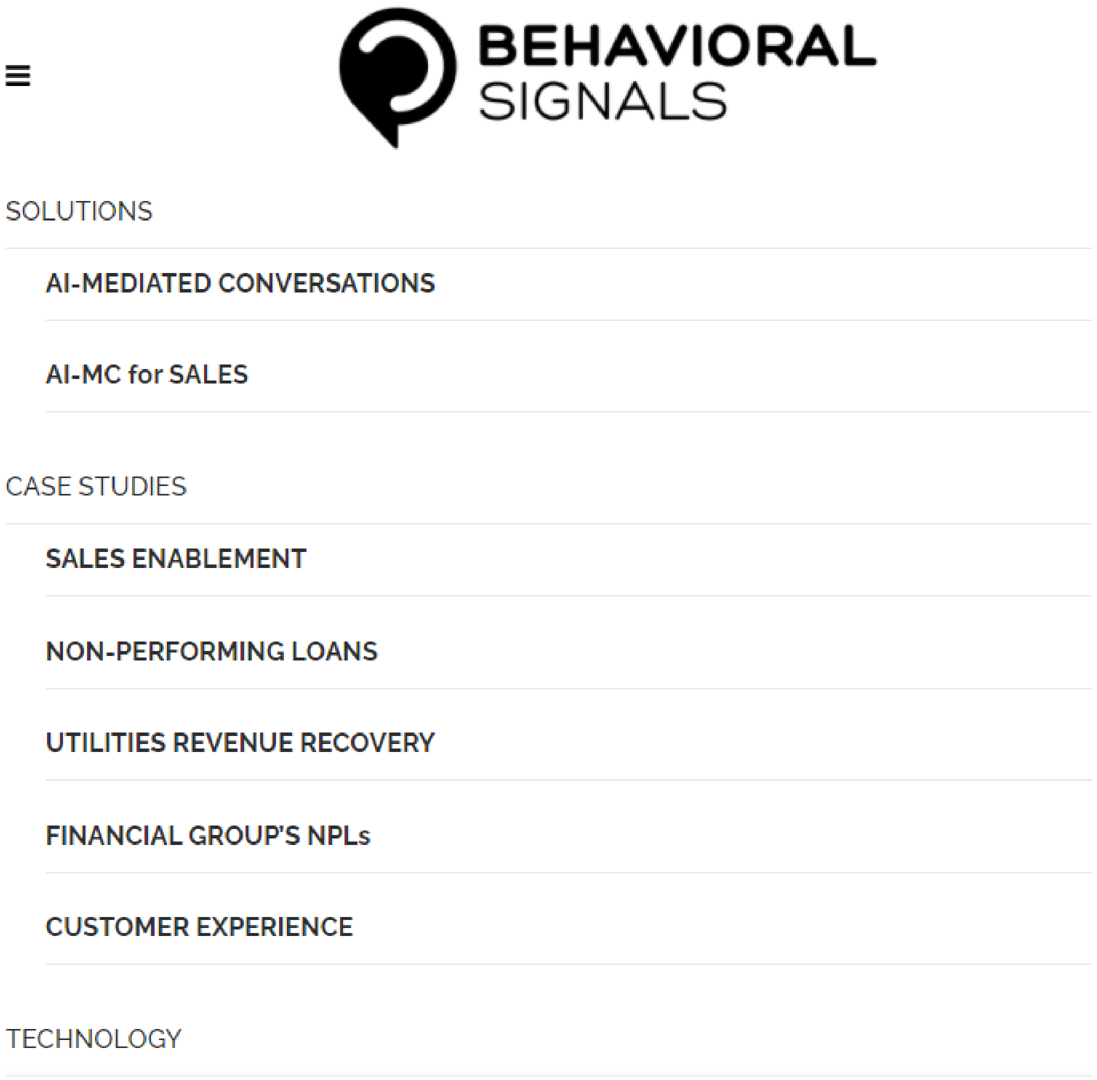
2) Siri

It iss am virtual assistant that is a massive part of Apple Inc. It uses voice gesture based And it works on natural language based user interface and it does various things like :

1) recommendations
2)answers questions
3)performs actions

3) Behavioral signals

AI-Mediated Conversations (AI-MC) is an automated telecommunications solution that uses emotional AI and voice data to match the client and agent most suited to manage a particular phone. These similarities are based on profile data and our advanced algorithms developed from years of research and knowledge in NLP and behavioral signal processing. Whatever the purpose, there is always a catalyst that can allow both parties to achieve the desired result. That influential factor is usually the simplest and most naturalprocess: the interaction or interaction that takes place between people. No matter what the type of businessconnection (sales call, support, collection), there will always be interactions between real people, where it israre for the relationship to be the same between two pairs of people.



≡

BEHAVIORAL SIGNALS

SOLUTIONS

- AI-MEDIATED CONVERSATIONS
- AI-MC for SALES

CASE STUDIES

- SALES ENABLEMENT
- NON-PERFORMING LOANS
- UTILITIES REVENUE RECOVERY
- FINANCIAL GROUP'S NPLs
- CUSTOMER EXPERIENCE

TECHNOLOGY

Fig6.web page of behavioral signals

4) Watson natural language understanding

Watson Natural Language Understanding is a native cloud product that uses in-depth reading toextract metadata from text such as organizations, keywords, categories, feelings, emotions, relationships, and syntax. Navigate under the topics mentioned in your data by using textanalysis to extract keywords, concepts, categories and more. Analyze your random data inmore than thirteen languages. Out-of-the-box typing models provide a high level of accuracyin all your content. Use Watson's Natural Language Understanding behind your security phoneor on any cloud. Train Watson to understand the language of your business and extract custom information through Watson Knowledge Studio. Keep your data secure with the assurancethat your data is safe and secure. IBM will not collect or store your data. Through ouradvanced native language processing service (NLP), we provide developers with tools forprocessing and extracting important information from informal data.

6) Emozo

The Emyzo DIY SaaS Research & Feedback Collection uses ethical and emotional comprehension to help you make the right decisions for all digital content. The Emozo forum helps you go beyond the normal customer data analysis and examine the hearts and minds of customers to understand the functionality and impact of all digital content. You can use Emozo to test the effectiveness of ads, apps, streaming media content, and other favorites, on any channel - the web, mobile, social media, TV, etc. The Emozo novel's method of combining unconsciousness (attention and emotion) with verbal (survey) responses helps you to understand the functionality of all digital content very quickly. Emozo uses AI to enable quality and speed resea rch on customer devices. Emozo supports processes to improve duplicate design and provides fully securedata protection for you and your customers.

1. Objective of Major Project:

The goal of this project is to give entire access to hands free control world, a certain boon To disabled humans and for people who can’t handle and use technology as well as the young generation does. Even the earliest researches on the topic were that of medical decitation software. It can be a certain help for students who are visually impaired and requireConstant assistance as they may find working on laptops and computers impossible such Technology can provide them an ease and comfort to move towards an easy lifestyle.

3. Use Case Diagram of the Major Project

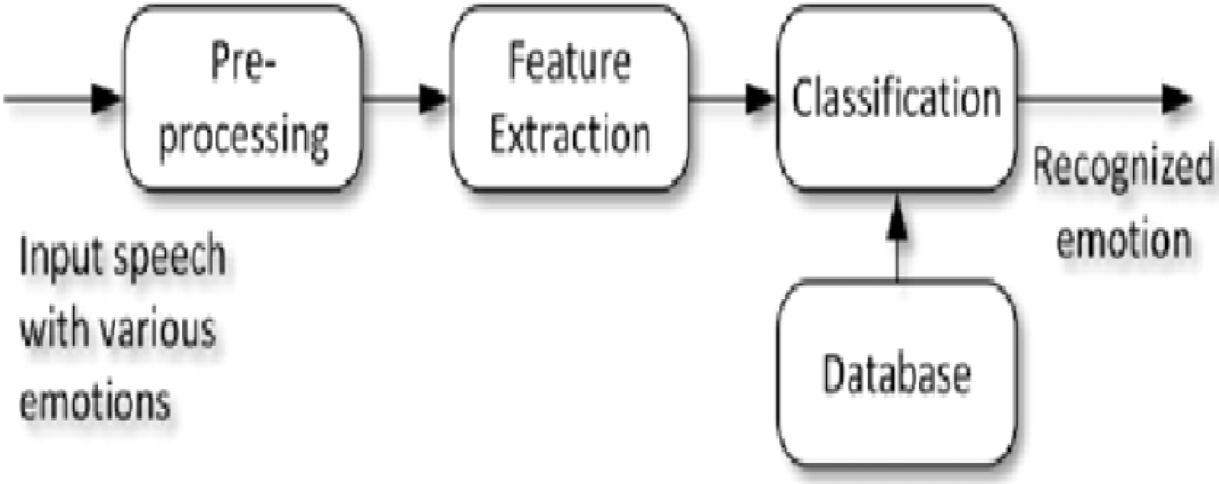


Fig 7. Use case Diagram

CHAPTER 3

System Development

3. Implementation of the Major Project

3.1 Technical Requirements (Software)

1)Jupyter Notebook

Jupyter notebooks basically provide an interactive computer space for developing

Python applications for data science. They were formerly known as ipthon testbo Oks. The following are some of the features of jupyter notebook that makes itt on e of the best part of python ML ecosystem:

1. Jupyter notebooks can show the process of step-by-step analysis by editing Things like code, output, text, images and everything. In a step-by-step manner.

2. Helps a data scientist write a thought process while developing an analysis Process.

3.One can also capture the result as part of a brochure.

4.With the help of jupyter notebook, we can share our work with our peers asWell.

2)Anaconda

Anaconda distribution contains conda and anaconda navigator, as well as python and Hundreds of science packages. When you install anaconda,, you also get all these. Conda works on the interface of your command line like anaconda prompt for wind Ows and terminals on mac0s and linux. Navigator is a desktop user image that lets you Launch apps and easily manages code packages, locations and channles without having To use command line commands.
It is used in several branches such as data science, machine learning, deep learning, etc.
It holds more than 200 libraries

Visual code studio
Google chrome
Css(bootstrap)
Css(external css)
Javascript
python

3.2 Dataset Used in the Major Project

We used the RAVDESS dataset for this machine learning project; RAVDESS stands for Ryerson Audio-Visual Database of Emotional Speech and Song, and it's available for free download. This database contains 7356 files that have been evaluated 247 times out of 10 in terms of emotional, dynamic and realistic performance.The entire database is 24.8GB from 24 characters.
In our project we have used a dataset consisting of 2800 files.

1.)

At first the loading of dataset is performed using the pandas library which allows to load

ll the 2800 files into our program, so that it can be used further for evaluation.

2.)

After that those datasets are converted from audio files to text files using the speechemotion API.

3.)

Then masking task is performed so that the unwanted noise which comes with the dataset is removed in order to perform perfect result. It is also known as the cleaning

process of the project.

4)

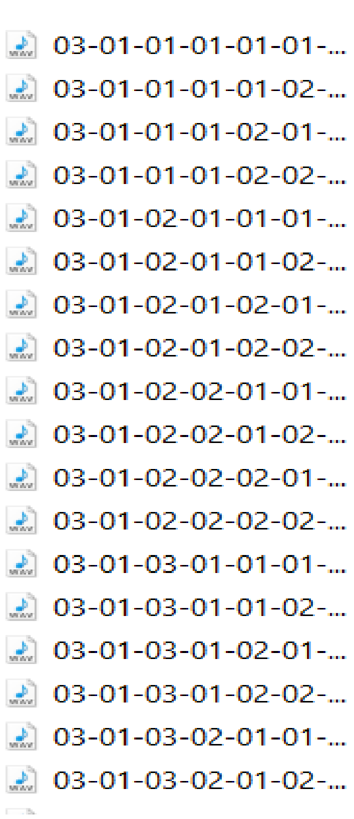
After the masking task comes the formation of graphs or we can say the spectrogramsusing the numpy library and the matplotlib.The matplotlib is the plotting library which is used to plot graphs in python and itsmathematical extensions comes under numpy.**.rolling** function is used for signal processing.The graph is plotted between the sound amplitude and the time where time in seconds

(s) which is further used for evaluation.

Here various graphs are plotted as there are 2800 files, we have shown some of them in the screenshots at the further.

5)

After this, depth visualization of the audio files is performed which refertothefeature extraction process, where various graphs are again plotted so among the different features of the dataset which are extracted from the previously implemented graphs. Then the dataset is again loaded and plots are visualized by calling the plotting function.



03-01-01-01-01-01-...
03-01-01-01-01-02-...
03-01-01-01-02-01-...
03-01-01-01-02-02-...
03-01-02-01-01-01-...
03-01-02-01-01-02-...
03-01-02-01-02-01-...
03-01-02-01-02-02-...
03-01-02-02-01-01-...
03-01-02-02-01-02-...
03-01-02-02-02-01-...
03-01-02-02-02-02-...
03-01-03-01-01-01-...
03-01-03-01-01-02-...
03-01-03-01-02-01-...
03-01-03-01-02-02-...
03-01-03-02-01-01-...
03-01-03-02-01-02-...

Fig.8 Dataset used

3.3 Approach of the Project Problem

1.Computational analysis

The whims of human language have made development difficult. It is considered one of the most complex areas of computer science, such as linguistics, mathematics and statistics. The speech recognition function consists of several components such as speech input, feature extraction, feature vector, decoder, and word output. The decoder uses an acoustic model, a pronunciation dictionary, and a language model to determine the appropriate output. Speech recognition technology is evaluated based on its accuracy rate. H. Word error rate (WER), and speed rating.

There are several factors that can affect the word error rate. B. Pronunciation, accents, pitch, volume, and background noise. Achieving human equivalence (the same error rate that two people speak) has long been a goal of speech recognition systems.

Various algorithms and computational techniques are used to convert speech to text and improve transcription accuracy. Below is a brief description of some of the most commonly used methods:

Natural language processing:

NLP is not necessarily a specific algorithm used for speech recognition, but a branch of artificial intelligence that focuses on the interaction between humans and machines through language through speech and text. Many mobile devices include voice recognition in their systems to perform voice searches. Siri - or provides additional options for texting.

Applications of natural language processing:

a) Sentiment Analysis

b) Text Classification

c) Chatbots & Virtual Assistants

d) Text Extraction

e) Machine Translation

f) Text Summarization

g) Market Intelligence

h) Auto-Correct

i) Intent Classification

j) Urgency Detection

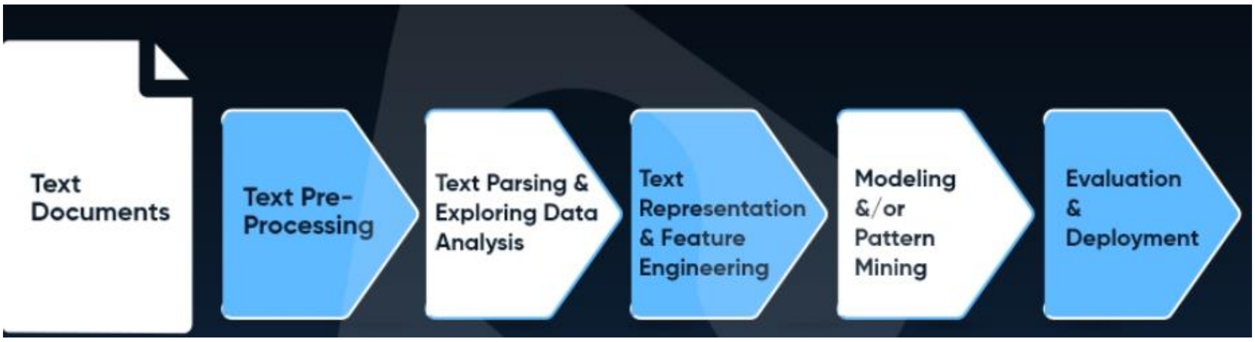

Text Documents
Text Pre-Processing
Text Parsing & Exploring Data Analysis
Text Representation & Feature Engineering
Modeling &/or Pattern Mining
Evaluation & Deployment

Fig 9. How natural language processing works

Hidden markov models :

The hidden Markov model is based on the Markov chain model, which assumes that the probability of a given state depends on the current state rather than the previous one. The Markov chain model is useful for observable events, such as text input, but the hidden Markov model allows you to include hidden events, such as part-of-speech tags, into the probabilistic model. They are used as sequence models in speech recognition to assign labels to each unit. Words, syllables, sentences, etc.

in order. These labels allow you to determine the most appropriate label order by generating a match for the given input.

N-gram :

This is the simplest type of language model (LM) that assigns probabilities to sentences or phrases. An Ngram is a sequence of Nwords. For example, "Order Pizza" is trigrams or 3 grams, and "Order Pizza" is 4 grams. Improve recognition and accuracy by using the grammar and probability of specific word sequences.

Neural networks:

Primarily leveraged for deep learning algorithms, neural networks process training data by mimicking the interconnectivity of the human brain through layers of nodes. Each node is made up of inputs, weights, a bias (or threshold) and an output. If that output value exceeds a given threshold, it “fires” or activates the node, passing data to the next layer in the network. Neural networks learn this mapping function through supervised learning, adjusting based on the loss function through the process of gradient descent. While neural networks tend to be more accurate and can accept more data, this comes at a performance efficiency cost as they tend to be slower to train compared to traditional language models.

Application neural network:

a) Artificial Neural Network (ANN)

b) Facial Recognition.

c) Stock Market Prediction.

d) Social Media.

e) Aerospace.

f) Defence.

g) Healthcare.

h) Signature Verification and Handwriting Analysis.

2. Used cases for SER(speech emotion recognition)

A wide number of industries are utilizing different applications of speech technology today, helping businesses and consumers save time and even lives. Some examples include:

1)Automotive:

Speech recognizers improves driver safety by enabling voiceactivated navigation systems and search capabilities in car radios.

2)Technology:

Virtual assistants are increasingly becoming integrated within our daily lives, particularly on our mobile devices. We use voice commands to access them through our smartphones, such as through Google Assistant or Apple’s Siri, for tasks, such as voice search, or through our speakers, via Amazon’s Alexa or Microsoft’s Cortana, to play music. They’ll only continue to integrate into the everyday products that we use, fueling the “Internet of Things” movement.

3)Healthcare:

Doctors and nurses leverage dictation applications to capture and log patient diagnoses and treatment notes.

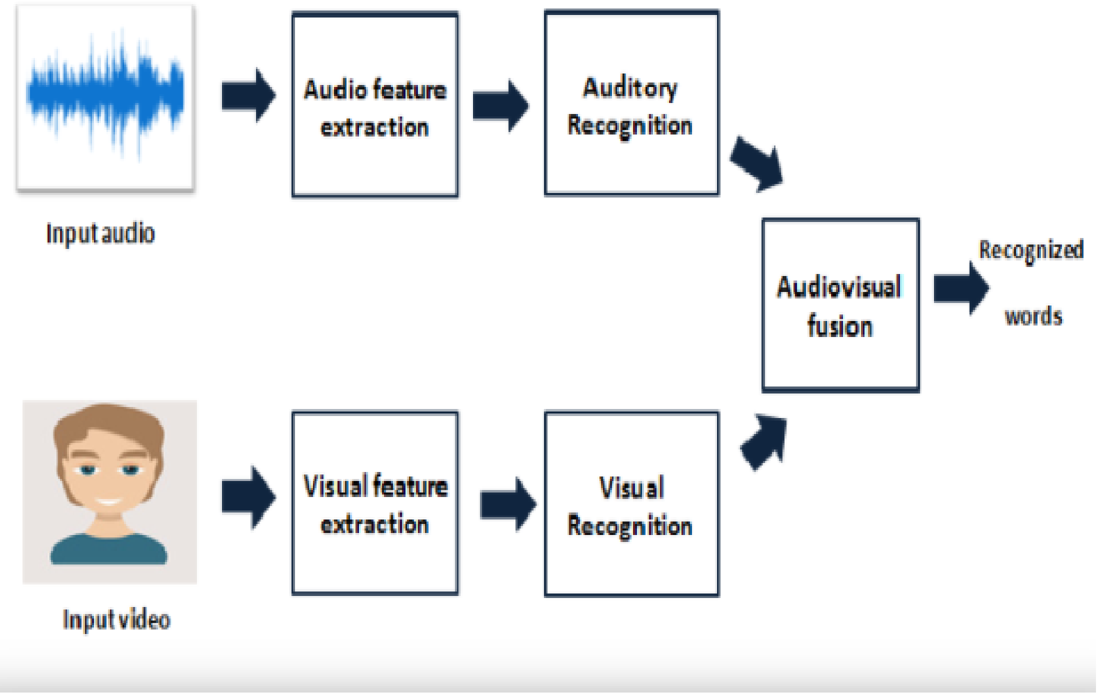
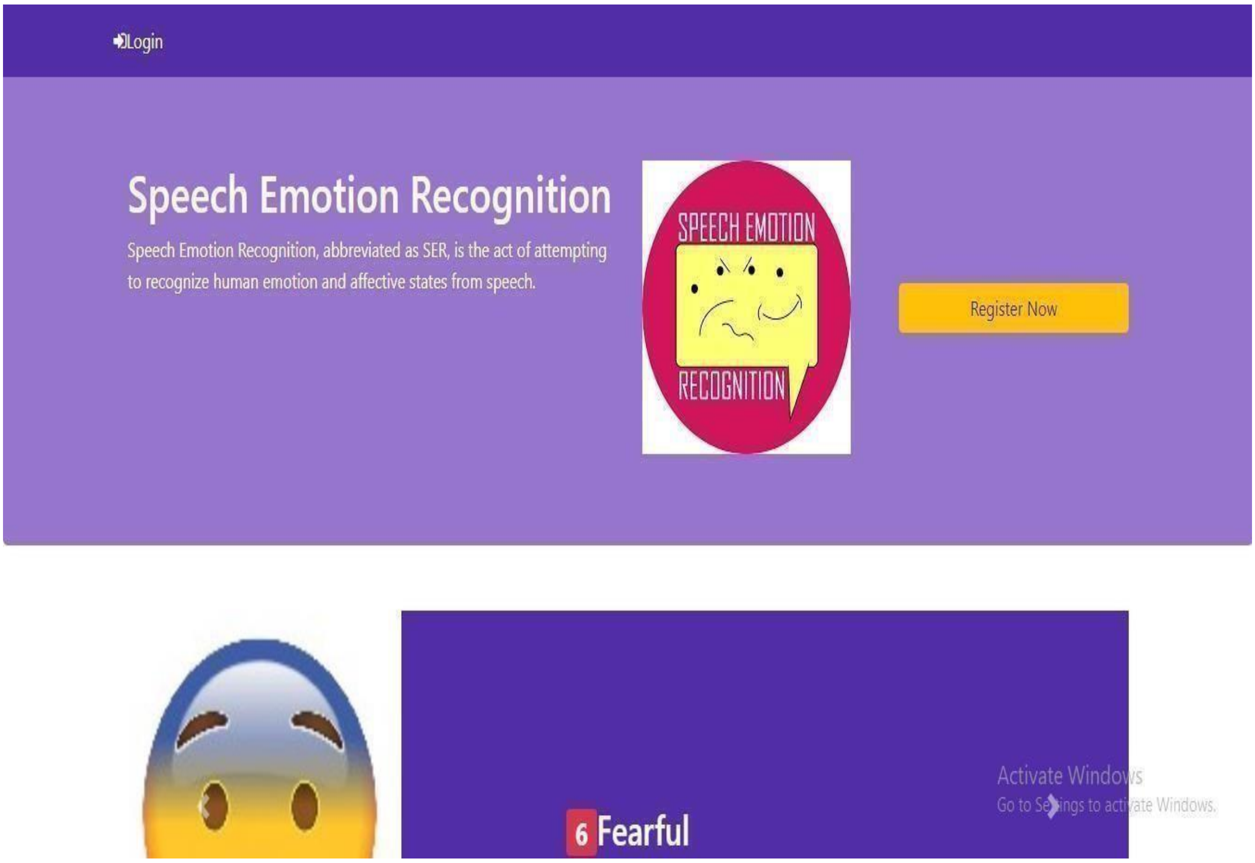


Fig.10 Model for speech recognition system

2. User interface



Speech Emotion Recognition

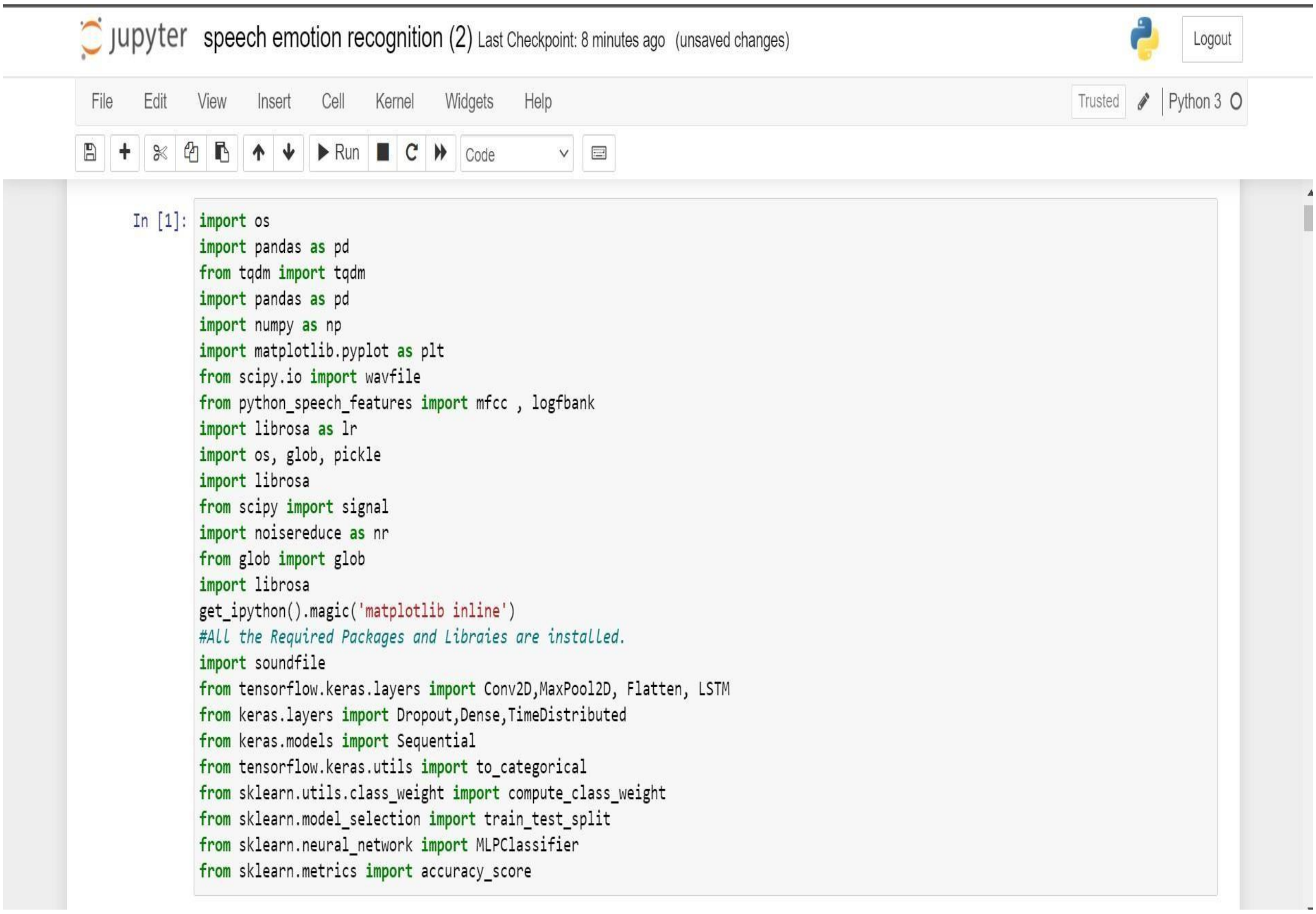
Speech Emotion Recognition, abbreviated as SER, is the act of attempting to recognize human emotion and affective states from speech.

Fig.11 Webpage of the project

3. Various stages of the project

Step 1

Importing the required libraries to read and run the input we provide.

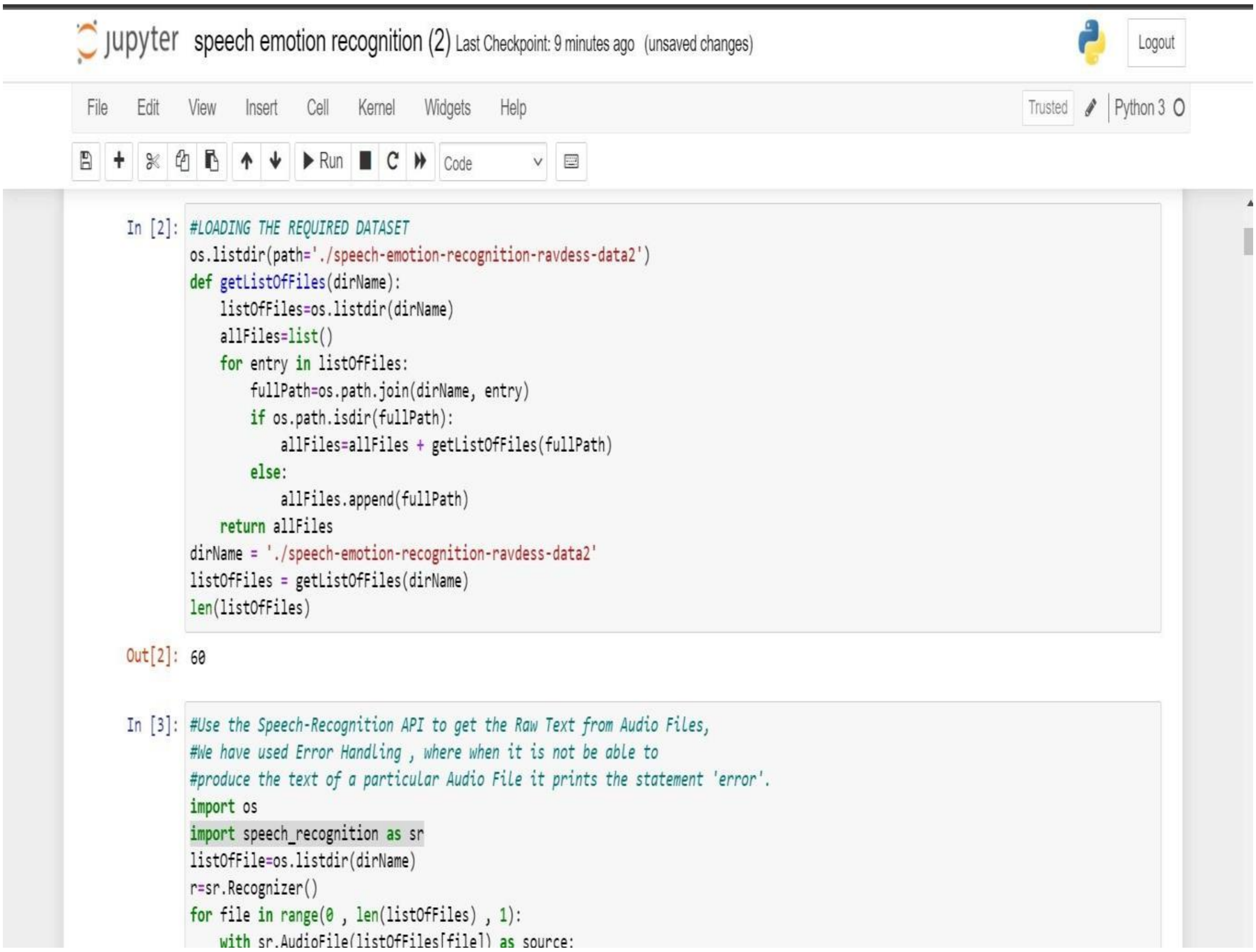


jupyter
speech emotion recognition (2) Last Checkpoint: 8 minutes ago (unsaved changes)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3
Run
Code

```
In [1]: import os
        import pandas as pd
        from tqdm import tqdm
        import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        from scipy.io import wavfile
        from python_speech_features import mfcc , logfbank
        import librosa as lr
        import os, glob, pickle
        import librosa
        from scipy import signal
        import noisereduce as nr
        from glob import glob
        import librosa
        get_ipython().magic('matplotlib inline')
        #ALL the Required Packages and Libraies are installed.
        import soundfile
        from tensorflow.keras.layers import Conv2D,MaxPool2D, Flatten, LSTM
        from keras.layers import Dropout,Dense,TimeDistributed
        from keras.models import Sequential
        from tensorflow.keras.utils import to_categorical
        from sklearn.utils.class_weight import compute_class_weight
        from sklearn.model_selection import train_test_split
        from sklearn.neural_network import MLPClassifier
        from sklearn.metrics import accuracy_score
```

Fig.12 Importing libraries

Step2


jupyter speech emotion recognition (2) Last Checkpoint: 9 minutes ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code

In [2]:
```
#LOADING THE REQUIRED DATASET
os.listdir(path='./speech-emotion-recognition-ravdess-data2')
def getListOfFiles(dirName):
    listOfFiles=os.listdir(dirName)
    allFiles=list()
    for entry in listOfFiles:
        fullPath=os.path.join(dirName, entry)
        if os.path.isdir(fullPath):
            allFiles=allFiles + getListOfFiles(fullPath)
        else:
            allFiles.append(fullPath)
    return allFiles
dirName = './speech-emotion-recognition-ravdess-data2'
listOfFiles = getListOfFiles(dirName)
len(listOfFiles)
```

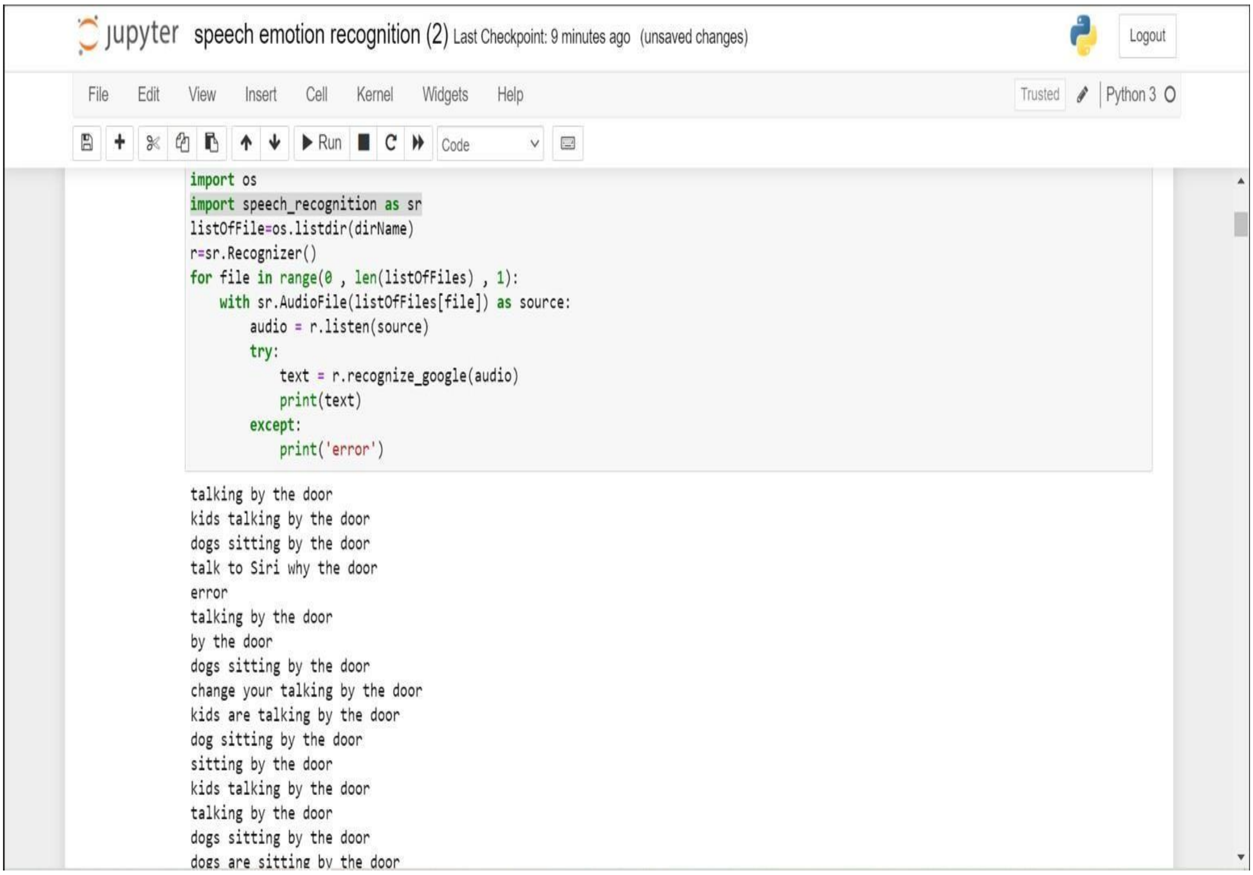
Out[2]:
```
60
```

In [3]:
```
#Use the Speech-Recognition API to get the Raw Text from Audio Files,
#We have used Error Handling , where when it is not be able to
#produce the text of a particular Audio File it prints the statement 'error'.
import os
import speech_recognition as sr
listOfFile=os.listdir(dirName)
r=sr.Recognizer()
for file in range(0 , len(listOfFiles) , 1):
    with sr.AudioFile(listOfFiles[file]) as source:
```

Fig.13 Loading of RAVEDESS dataset

Step 3

In order to utilize our dataset and work on our project we convert audio files into text file.


jupyter
speech emotion recognition (2) Last Checkpoint: 9 minutes ago (unsaved changes)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3
Run
Code

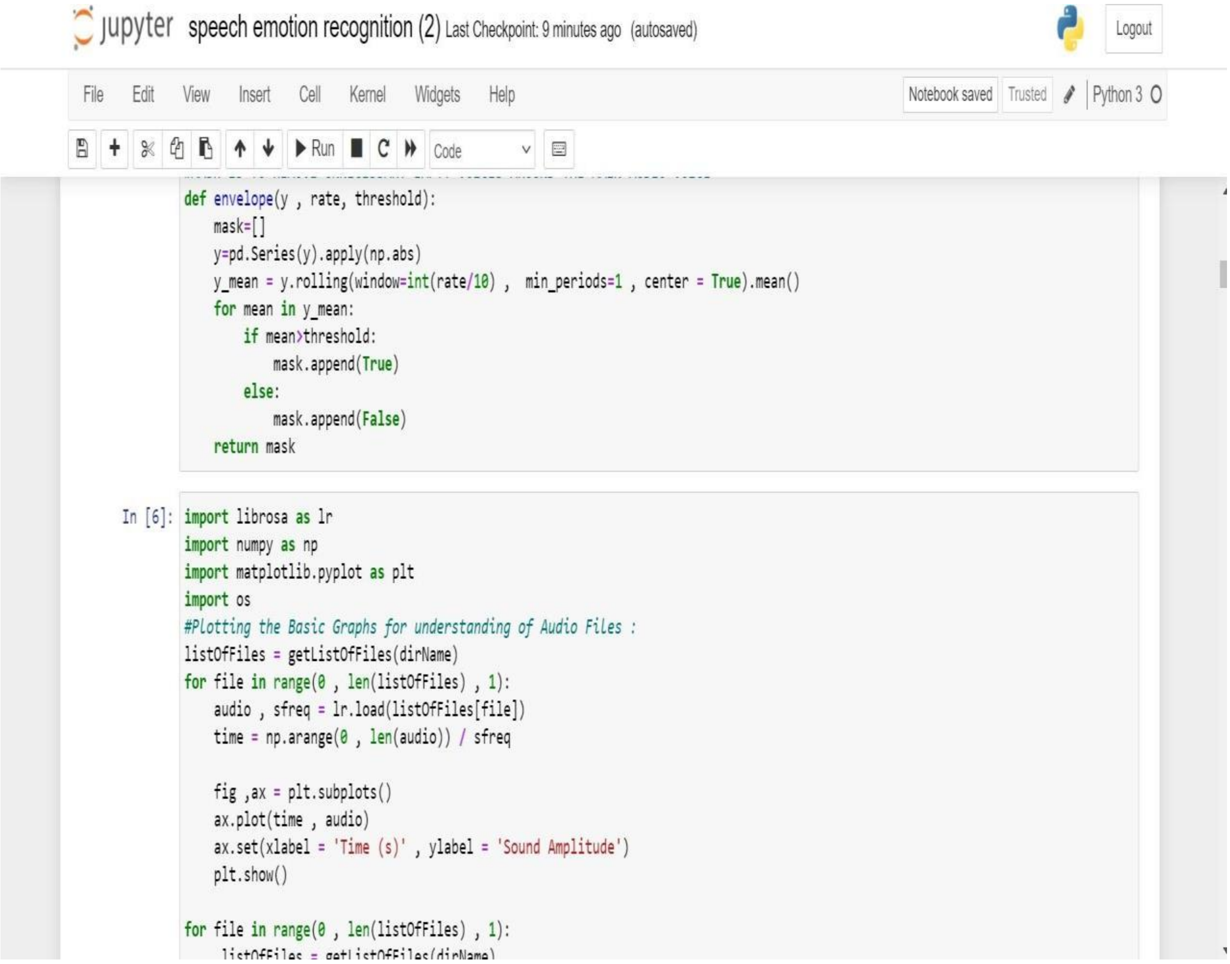
```
import os
import speech_recognition as sr
listOfFile=os.listdir(dirName)
r=sr.Recognizer()
for file in range(0 , len(listOfFiles) , 1):
    with sr.AudioFile(listOfFiles[file]) as source:
        audio = r.listen(source)
        try:
            text = r.recognize_google(audio)
            print(text)
        except:
            print('error')
```

```
talking by the door
kids talking by the door
dogs sitting by the door
talk to Siri why the door
error
talking by the door
by the door
dogs sitting by the door
change your talking by the door
kids are talking by the door
dog sitting by the door
sitting by the door
kids talking by the door
talking by the door
dogs sitting by the door
dogs are sitting by the door
```

Fig.13 conversion of audio files into text files

Step 4

We process our data by removing all the unwanted data in our dataset.



jupyter speech emotion recognition (2) Last Checkpoint: 9 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Notebook saved Trusted Python 3
Run Code

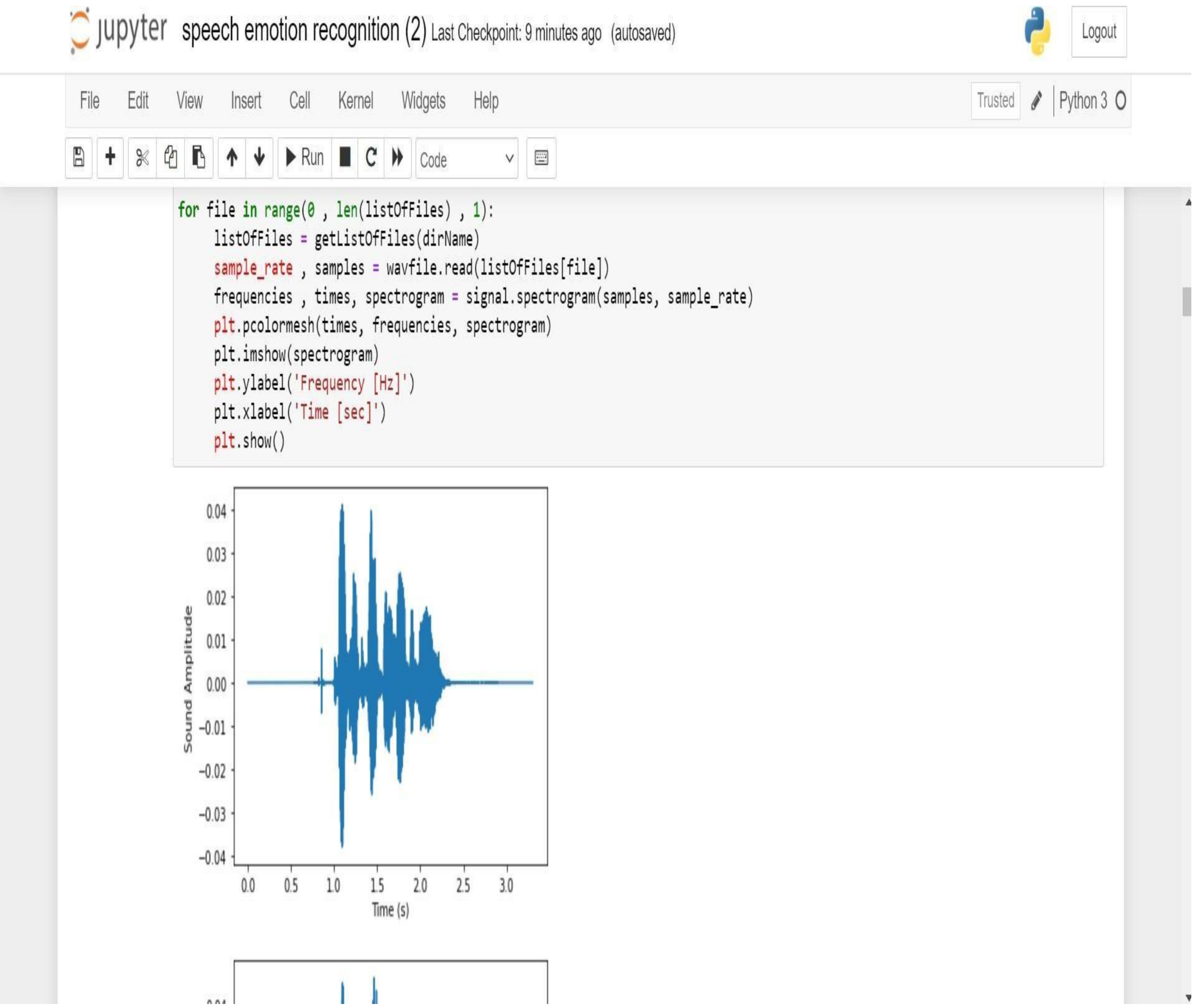
```
def envelope(y , rate, threshold):
    mask=[]
    y=pd.Series(y).apply(np.abs)
    y_mean = y.rolling(window=int(rate/10) ,  min_periods=1 , center = True).mean()
    for mean in y_mean:
        if mean>threshold:
            mask.append(True)
        else:
            mask.append(False)
    return mask
```

```
In [6]: import librosa as lr
import numpy as np
import matplotlib.pyplot as plt
import os
#Plotting the Basic Graphs for understanding of Audio Files :
listOfFiles = getListOfFiles(dirName)
for file in range(0 , len(listOfFiles) , 1):
    audio , sfreq = lr.load(listOfFiles[file])
    time = np.arange(0 , len(audio)) / sfreq

    fig ,ax = plt.subplots()
    ax.plot(time , audio)
    ax.set(xlabel = 'Time (s)' , ylabel = 'Sound Amplitude')
    plt.show()

for file in range(0 , len(listOfFiles) , 1):
    listOfFiles = getListOfFiles(dirName)
```

Fig.14 Masking(removal of unwanted data)

Step 5

Analyzing the dataset through graphs.



```
for file in range(0 , len(listOfFiles) , 1):
    listOfFiles = getListOfFiles(dirName)
    sample_rate , samples = wavfile.read(listOfFiles[file])
    frequencies , times, spectrogram = signal.spectrogram(samples, sample_rate)
    plt.pcolormesh(times, frequencies, spectrogram)
    plt.imshow(spectrogram)
    plt.ylabel('Frequency [Hz]')
    plt.xlabel('Time [sec]')
    plt.show()
```

Fig.15 Sound amplitude vs time graph

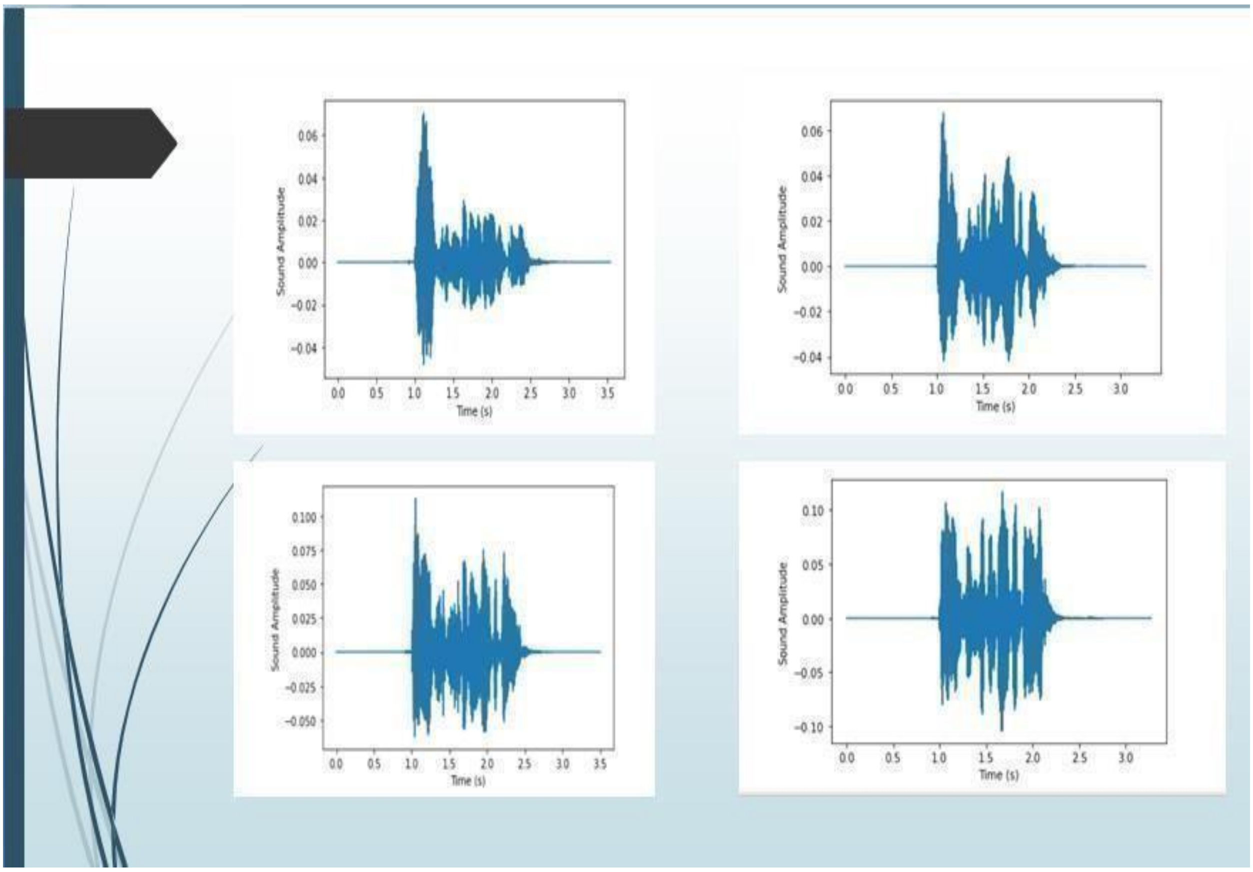


Fig.16 Sound amplitude vs time graph

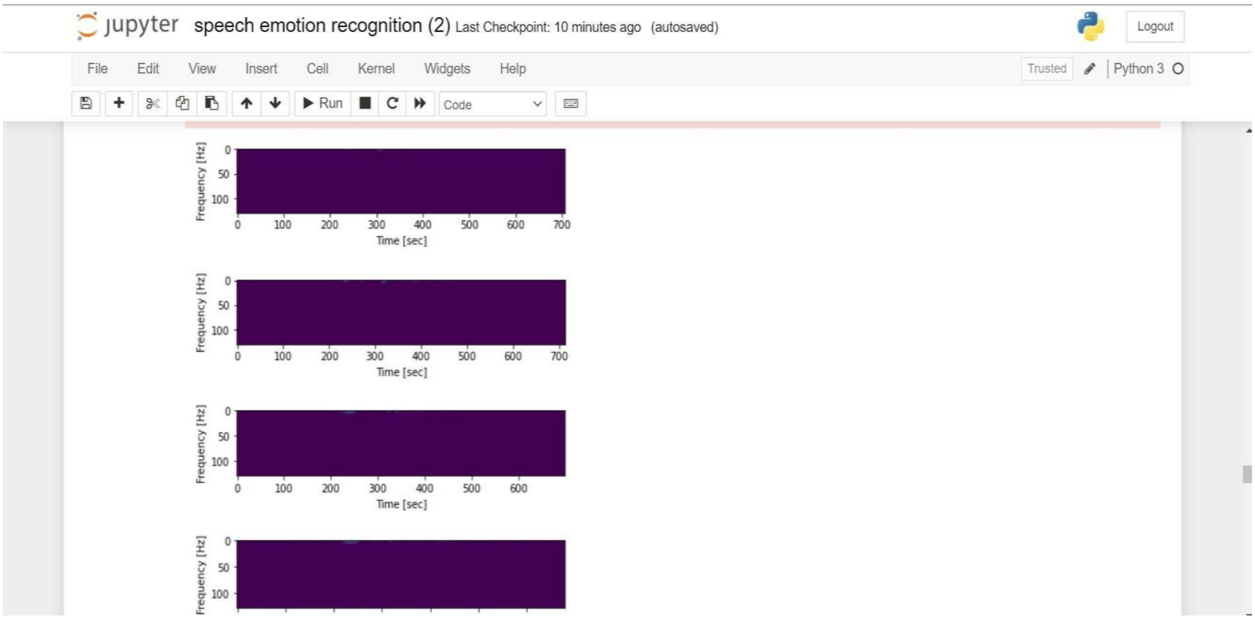
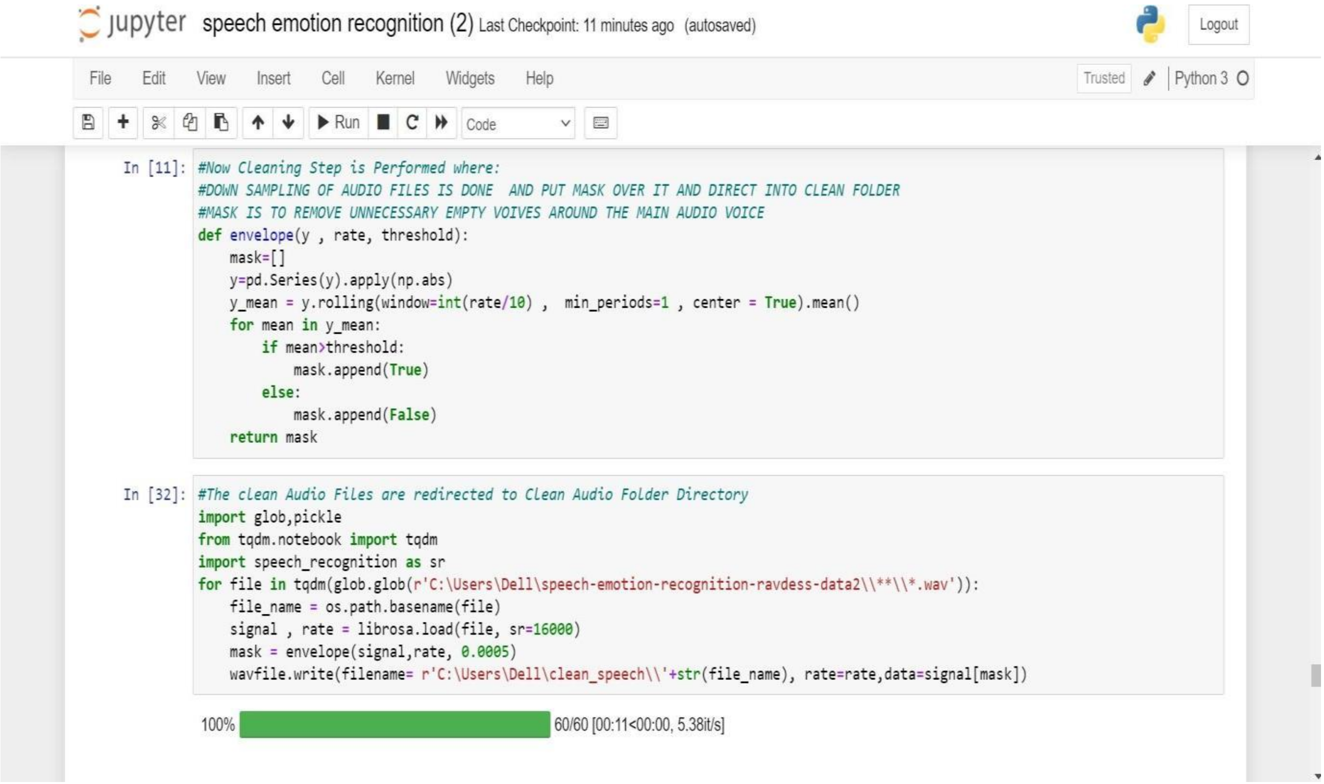


Fig.17 Frequency vs Time graph

Step 6

After pre-processing and filtering the audio files we again mask them and store them for further use.


jupyter speech emotion recognition (2) Last Checkpoint: 11 minutes ago (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3
Run
Code

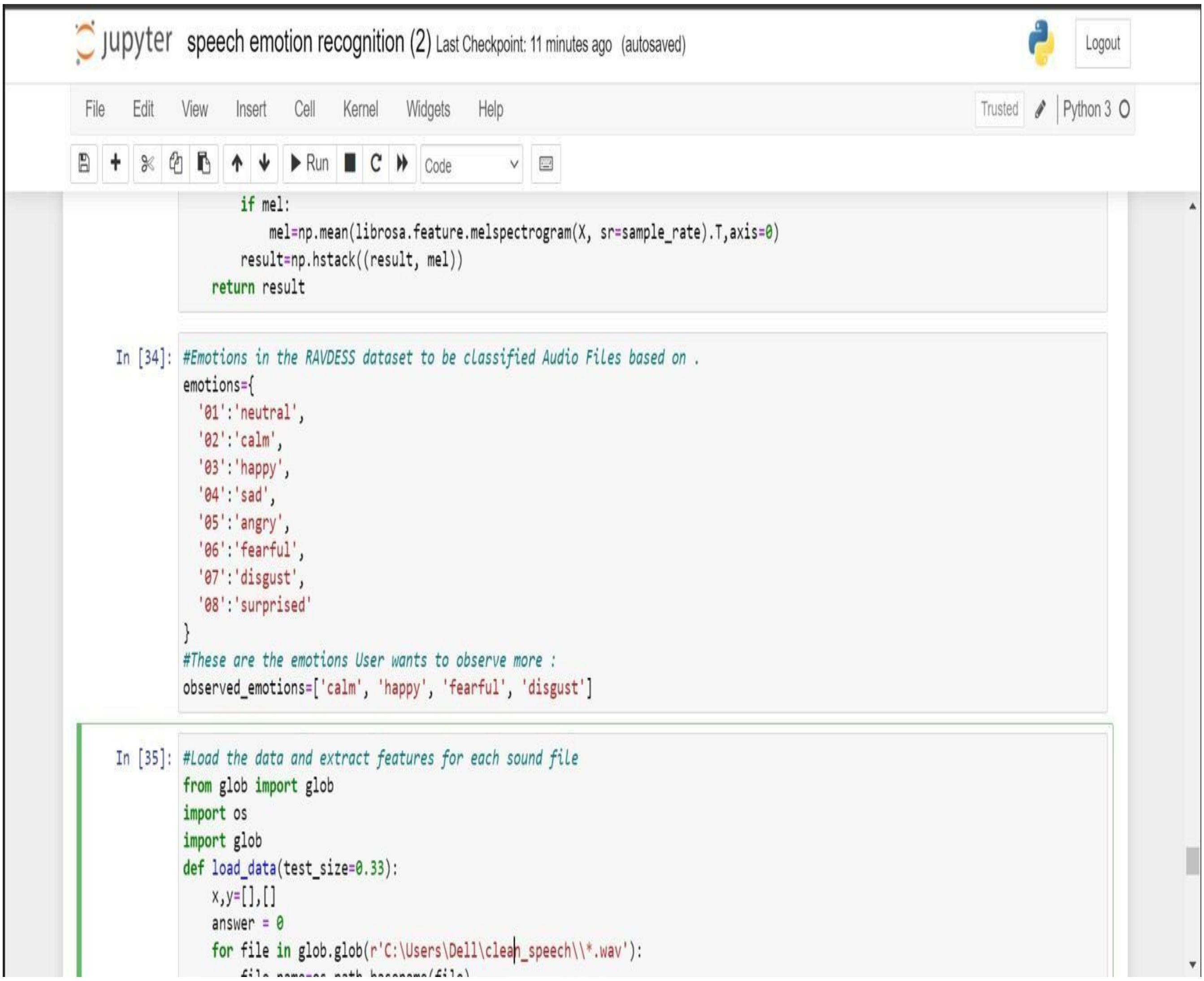
```
In [11]: #Now Cleaning Step is Performed where:
         #DOWN SAMPLING OF AUDIO FILES IS DONE  AND PUT MASK OVER IT AND DIRECT INTO CLEAN FOLDER
         #MASK IS TO REMOVE UNNECESSARY EMPTY VOIVES AROUND THE MAIN AUDIO VOICE
         def envelope(y , rate, threshold):
             mask=[]
             y=pd.Series(y).apply(np.abs)
             y_mean = y.rolling(window=int(rate/10) ,  min_periods=1 , center = True).mean()
             for mean in y_mean:
                 if mean>threshold:
                     mask.append(True)
                 else:
                     mask.append(False)
             return mask
```

```
In [32]: #The clean Audio Files are redirected to Clean Audio Folder Directory
         import glob,pickle
         from tqdm.notebook import tqdm
         import speech_recognition as sr
         for file in tqdm(glob.glob(r'C:\Users\Dell\speech-emotion-recognition-ravdess-data2\\**\\*.wav')):
             file_name = os.path.basename(file)
             signal , rate = librosa.load(file, sr=16000)
             mask = envelope(signal,rate, 0.0005)
             wavfile.write(filename= r'C:\Users\Dell\clean_speech\\'+str(file_name), rate=rate,data=signal[mask])

         100% ████████████████ 60/60 [00:11<00:00, 5.38it/s]
```

Fig.18 Second masking and storing clean files

Step 7

Labeling the emotions


jupyter speech emotion recognition (2) Last Checkpoint: 11 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code

```
        if mel:
            mel=np.mean(librosa.feature.melspectrogram(X, sr=sample_rate).T,axis=0)
        result=np.hstack((result, mel))
    return result
```

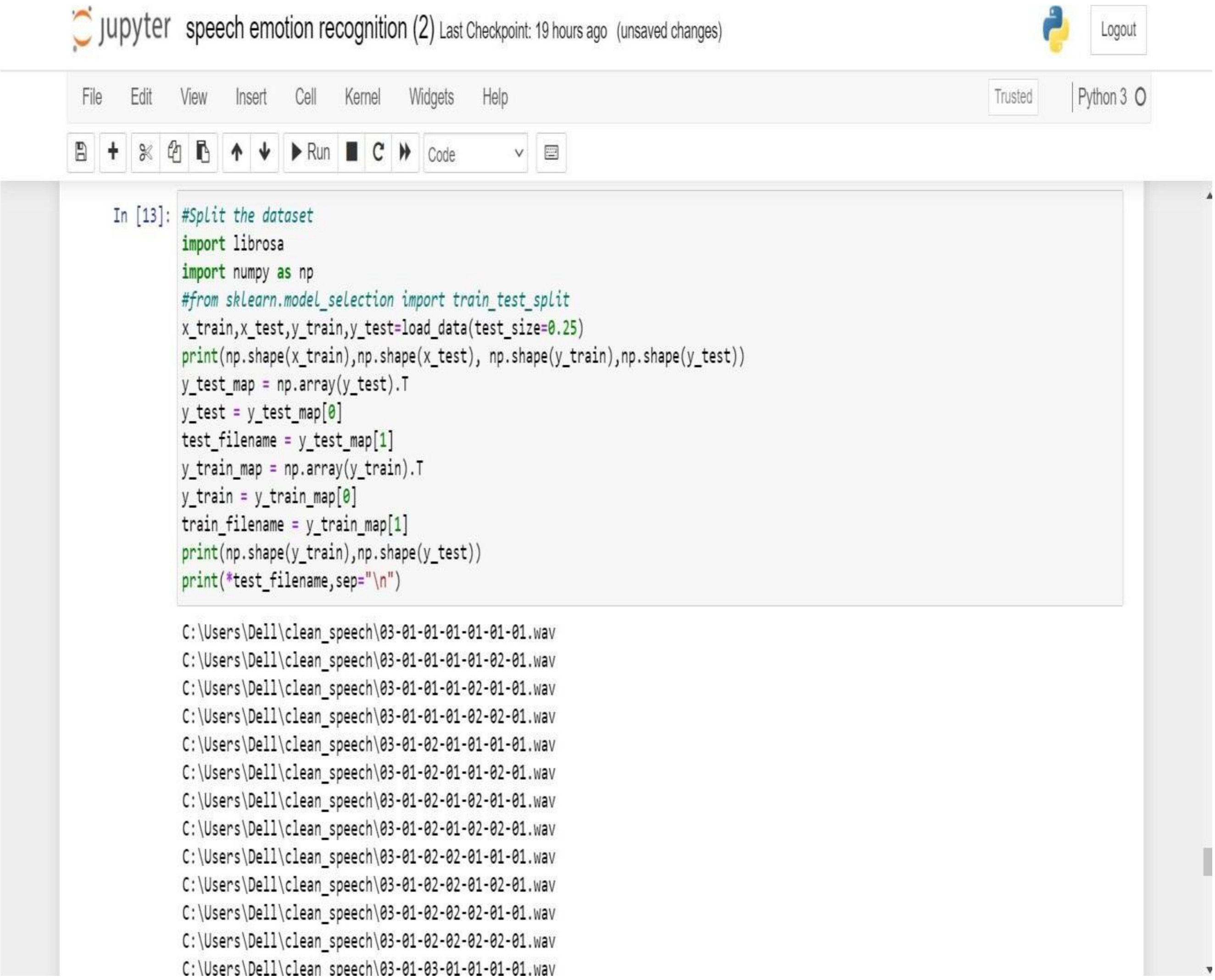
In [34]:
```
#Emotions in the RAVDESS dataset to be classified Audio Files based on .
emotions={
  '01':'neutral',
  '02':'calm',
  '03':'happy',
  '04':'sad',
  '05':'angry',
  '06':'fearful',
  '07':'disgust',
  '08':'surprised'
}
#These are the emotions User wants to observe more :
observed_emotions=['calm', 'happy', 'fearful', 'disgust']
```

In [35]:
```
#Load the data and extract features for each sound file
from glob import glob
import os
import glob
def load_data(test_size=0.33):
    x,y=[],[]
    answer = 0
    for file in glob.glob(r'C:\Users\Dell\clean_speech\\*.wav'):
        file_name=os.path.basename(file)
```

Fig.19 labeled emotions

Step 8

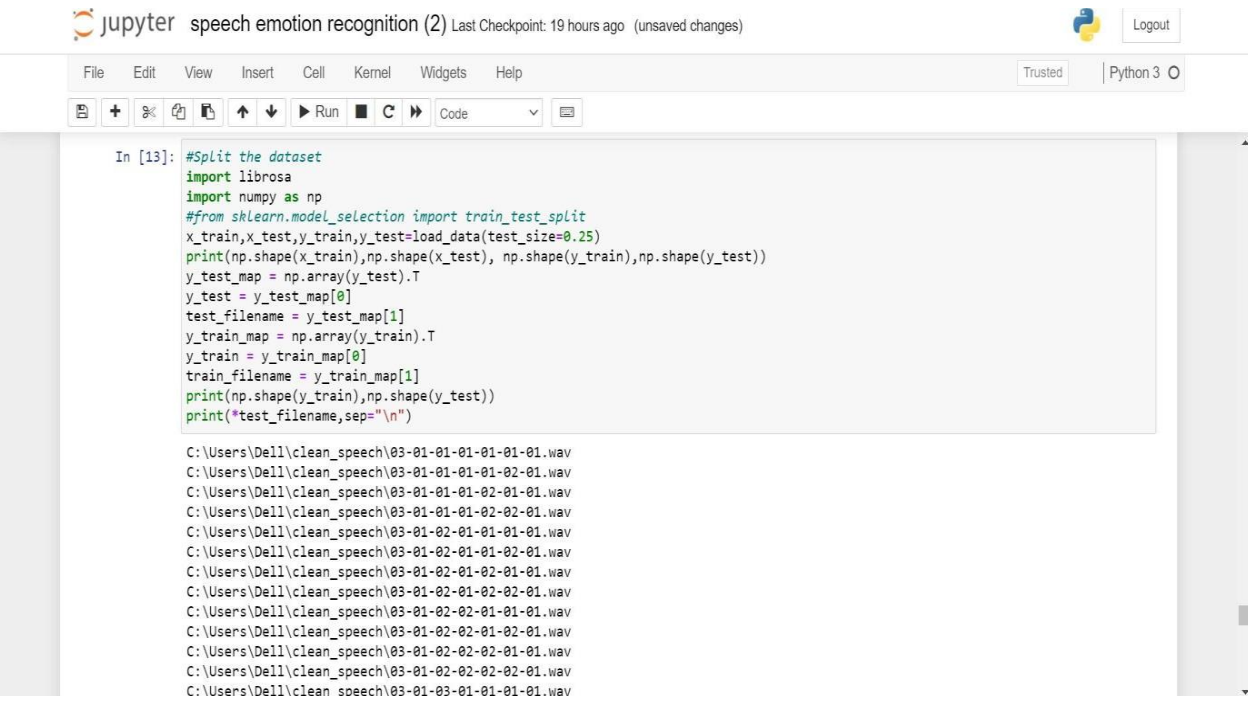
Importing the clean dataset that was stored earlier.


jupyter speech emotion recognition (2) Last Checkpoint: 19 hours ago (unsaved changes)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3
Run
Code

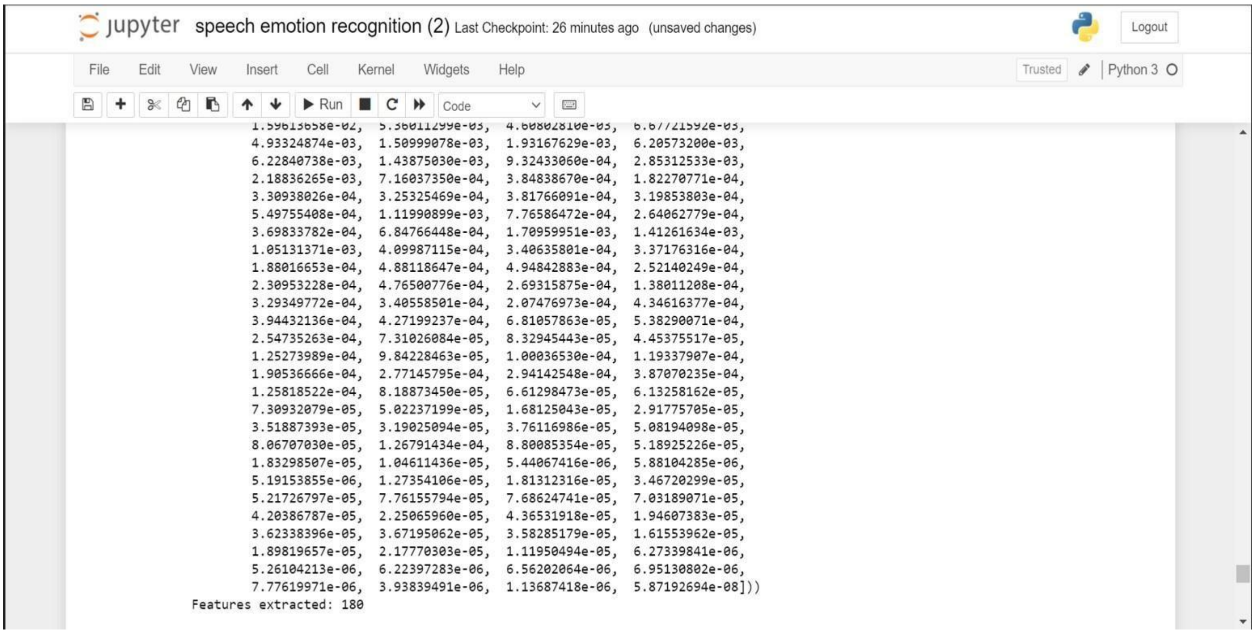
```
In [13]: #Split the dataset
         import librosa
         import numpy as np
         #from sklearn.model_selection import train_test_split
         x_train,x_test,y_train,y_test=load_data(test_size=0.25)
         print(np.shape(x_train),np.shape(x_test), np.shape(y_train),np.shape(y_test))
         y_test_map = np.array(y_test).T
         y_test = y_test_map[0]
         test_filename = y_test_map[1]
         y_train_map = np.array(y_train).T
         y_train = y_train_map[0]
         train_filename = y_train_map[1]
         print(np.shape(y_train),np.shape(y_test))
         print(*test_filename,sep="\n")
```

```
C:\Users\Dell\clean_speech\03-01-01-01-01-01-01.wav
C:\Users\Dell\clean_speech\03-01-01-01-01-02-01.wav
C:\Users\Dell\clean_speech\03-01-01-01-02-01-01.wav
C:\Users\Dell\clean_speech\03-01-01-01-02-02-01.wav
C:\Users\Dell\clean_speech\03-01-02-01-01-01-01.wav
C:\Users\Dell\clean_speech\03-01-02-01-01-02-01.wav
C:\Users\Dell\clean_speech\03-01-02-01-02-01-01.wav
C:\Users\Dell\clean_speech\03-01-02-01-02-02-01.wav
C:\Users\Dell\clean_speech\03-01-02-02-01-01-01.wav
C:\Users\Dell\clean_speech\03-01-02-02-01-02-01.wav
C:\Users\Dell\clean_speech\03-01-02-02-02-01-01.wav
C:\Users\Dell\clean_speech\03-01-02-02-02-02-01.wav
C:\Users\Dell\clean_speech\03-01-03-01-01-01-01.wav
```

Fig.20 Importing the cleaned dataset


jupyter
speech emotion recognition (2) Last Checkpoint: 19 hours ago (unsaved changes)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3
Run
Code

```
In [13]: #Split the dataset
         import librosa
         import numpy as np
         #from sklearn.model_selection import train_test_split
         x_train,x_test,y_train,y_test=load_data(test_size=0.25)
         print(np.shape(x_train),np.shape(x_test), np.shape(y_train),np.shape(y_test))
         y_test_map = np.array(y_test).T
         y_test = y_test_map[0]
         test_filename = y_test_map[1]
         y_train_map = np.array(y_train).T
         y_train = y_train_map[0]
         train_filename = y_train_map[1]
         print(np.shape(y_train),np.shape(y_test))
         print(*test_filename,sep="\n")

         C:\Users\Dell\clean_speech\03-01-01-01-01-01-01.wav
         C:\Users\Dell\clean_speech\03-01-01-01-01-02-01.wav
         C:\Users\Dell\clean_speech\03-01-01-01-02-01-01.wav
         C:\Users\Dell\clean_speech\03-01-01-01-02-02-01.wav
         C:\Users\Dell\clean_speech\03-01-02-01-01-01-01.wav
         C:\Users\Dell\clean_speech\03-01-02-01-01-02-01.wav
         C:\Users\Dell\clean_speech\03-01-02-01-02-01-01.wav
         C:\Users\Dell\clean_speech\03-01-02-01-02-02-01.wav
         C:\Users\Dell\clean_speech\03-01-02-02-01-01-01.wav
         C:\Users\Dell\clean_speech\03-01-02-02-01-02-01.wav
         C:\Users\Dell\clean_speech\03-01-02-02-02-01-01.wav
         C:\Users\Dell\clean_speech\03-01-02-02-02-02-01.wav
         C:\Users\Dell\clean_speech\03-01-03-01-01-01-01.wav
```

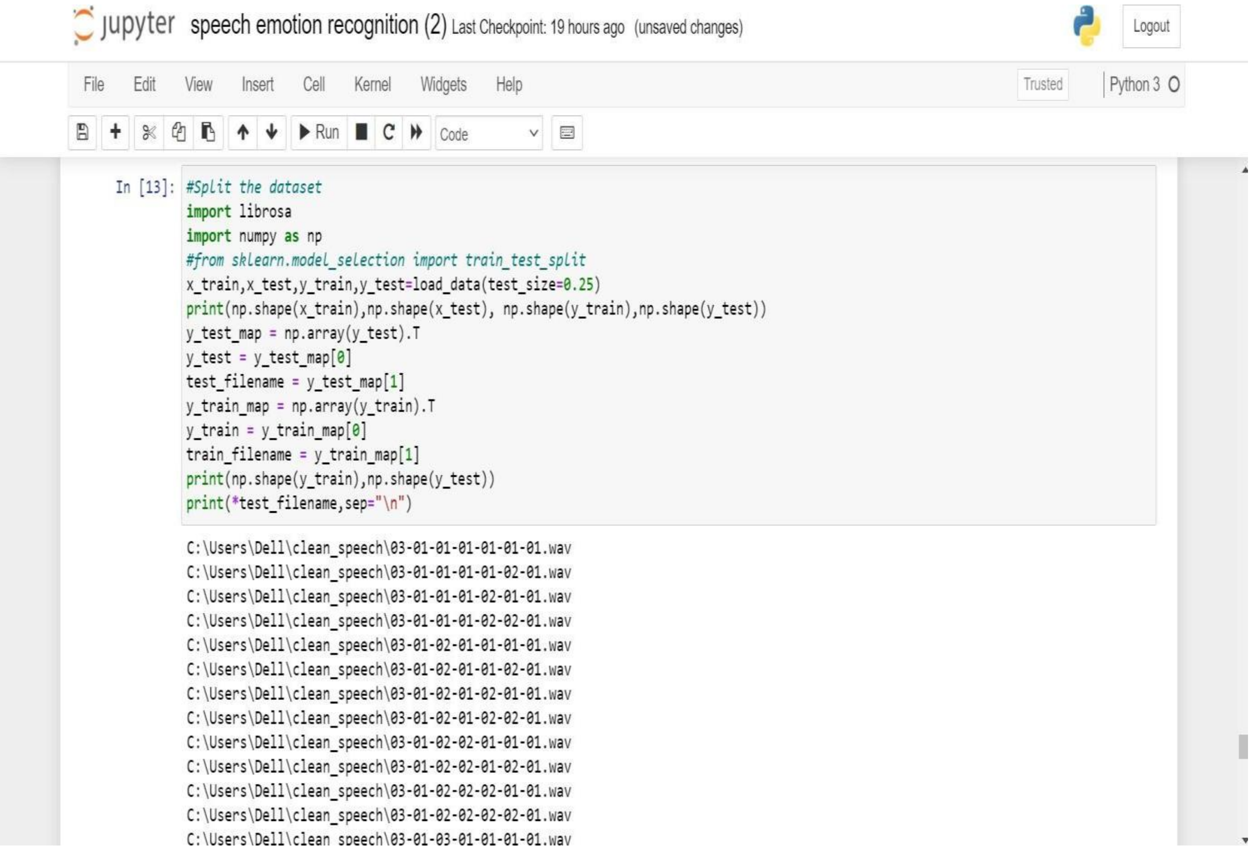

jupyter
speech emotion recognition (2) Last Checkpoint: 26 minutes ago (unsaved changes)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3
Run
Code

```
        1.59613658e-02, 5.36011299e-03, 4.60802810e-03, 6.67721592e-03,
        4.93324874e-03, 1.50999078e-03, 1.93167629e-03, 6.20573200e-03,
        6.22840738e-03, 1.43875030e-03, 9.32433060e-04, 2.85312533e-03,
        2.18836265e-03, 7.16037350e-04, 3.84838670e-04, 1.82270771e-04,
        3.30938026e-04, 3.25325469e-04, 3.81766091e-04, 3.19853803e-04,
        5.49755408e-04, 1.11990899e-03, 7.76586472e-04, 2.64062779e-04,
        3.69833782e-04, 6.84766448e-04, 1.70959951e-03, 1.41261634e-03,
        1.05131371e-03, 4.09987115e-04, 3.40635801e-04, 3.37176316e-04,
        1.88016653e-04, 4.88118647e-04, 4.94842883e-04, 2.52140249e-04,
        2.30953228e-04, 4.76500776e-04, 2.69315875e-04, 1.38011208e-04,
        3.29349772e-04, 3.40558501e-04, 2.07476973e-04, 4.34616377e-04,
        3.94432136e-04, 4.27199237e-04, 6.81057863e-05, 5.38290071e-04,
        2.54735263e-04, 7.31026084e-05, 8.32945443e-05, 4.45375517e-05,
        1.25273989e-04, 9.84228463e-05, 1.00036530e-04, 1.19337907e-04,
        1.90536666e-04, 2.77145795e-04, 2.94142548e-04, 3.87070235e-04,
        1.25818522e-04, 8.18873450e-05, 6.61298473e-05, 6.13258162e-05,
        7.30932079e-05, 5.02237199e-05, 1.68125043e-05, 2.91775705e-05,
        3.51887393e-05, 3.19025094e-05, 3.76116986e-05, 5.08194098e-05,
        8.06707030e-05, 1.26791434e-04, 8.80085354e-05, 5.18925226e-05,
        1.83298507e-05, 1.04611436e-05, 5.44067416e-06, 5.88104285e-06,
        5.19153855e-06, 1.27354106e-05, 1.81312316e-05, 3.46720299e-05,
        5.21726797e-05, 7.76155794e-05, 7.68624741e-05, 7.03189071e-05,
        4.20386787e-05, 2.25065960e-05, 4.36531918e-05, 1.94607383e-05,
        3.62338396e-05, 3.67195062e-05, 3.58285179e-05, 1.61553962e-05,
        1.89819657e-05, 2.17770303e-05, 1.11950494e-05, 6.27339841e-06,
        5.26104213e-06, 6.22397283e-06, 6.56202064e-06, 6.95130802e-06,
        7.77619971e-06, 3.93839491e-06, 1.13687418e-06, 5.87192694e-08]))
Features extracted: 180
```

Fig.21,22 total extracted features from the clean dataset

Step 9

Training the model with the help of multi-layer perception classifier(MLP classifier).


jupyter
speech emotion recognition (2) Last Checkpoint: 19 hours ago (unsaved changes)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3
Run
Code

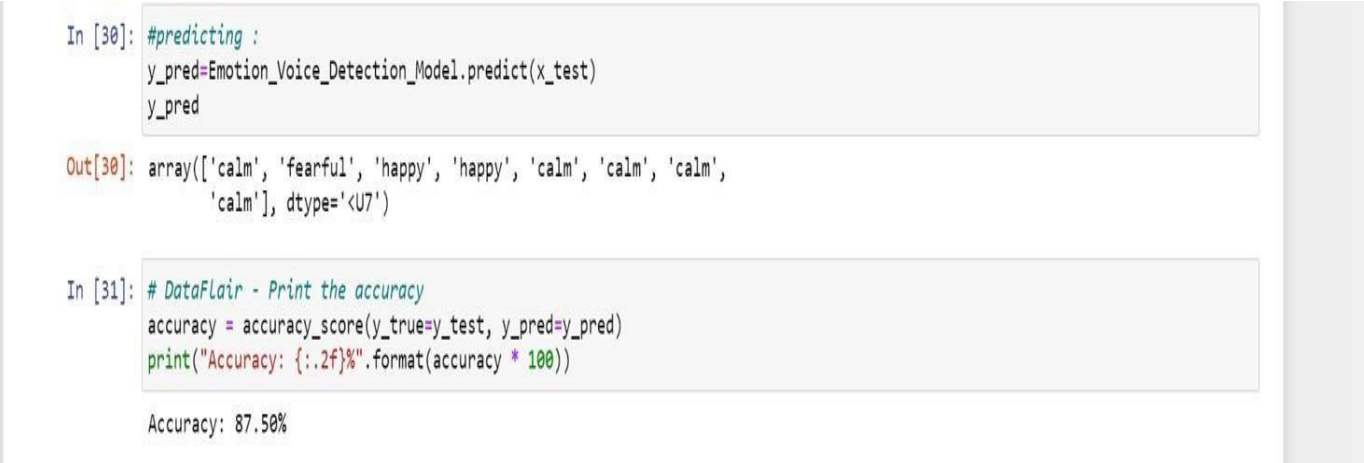
```
In [13]: #Split the dataset
import librosa
import numpy as np
#from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=load_data(test_size=0.25)
print(np.shape(x_train),np.shape(x_test), np.shape(y_train),np.shape(y_test))
y_test_map = np.array(y_test).T
y_test = y_test_map[0]
test_filename = y_test_map[1]
y_train_map = np.array(y_train).T
y_train = y_train_map[0]
train_filename = y_train_map[1]
print(np.shape(y_train),np.shape(y_test))
print(*test_filename,sep="\n")
```

```
C:\Users\Dell\clean_speech\03-01-01-01-01-01-01.wav
C:\Users\Dell\clean_speech\03-01-01-01-01-02-01.wav
C:\Users\Dell\clean_speech\03-01-01-01-02-01-01.wav
C:\Users\Dell\clean_speech\03-01-01-01-02-02-01.wav
C:\Users\Dell\clean_speech\03-01-02-01-01-01-01.wav
C:\Users\Dell\clean_speech\03-01-02-01-01-02-01.wav
C:\Users\Dell\clean_speech\03-01-02-01-02-01-01.wav
C:\Users\Dell\clean_speech\03-01-02-01-02-02-01.wav
C:\Users\Dell\clean_speech\03-01-02-02-01-01-01.wav
C:\Users\Dell\clean_speech\03-01-02-02-01-02-01.wav
C:\Users\Dell\clean_speech\03-01-02-02-02-01-01.wav
C:\Users\Dell\clean_speech\03-01-02-02-02-02-01.wav
C:\Users\Dell\clean_speech\03-01-03-01-01-01-01.wav
```

Fig.23 Initializing MLP classifier

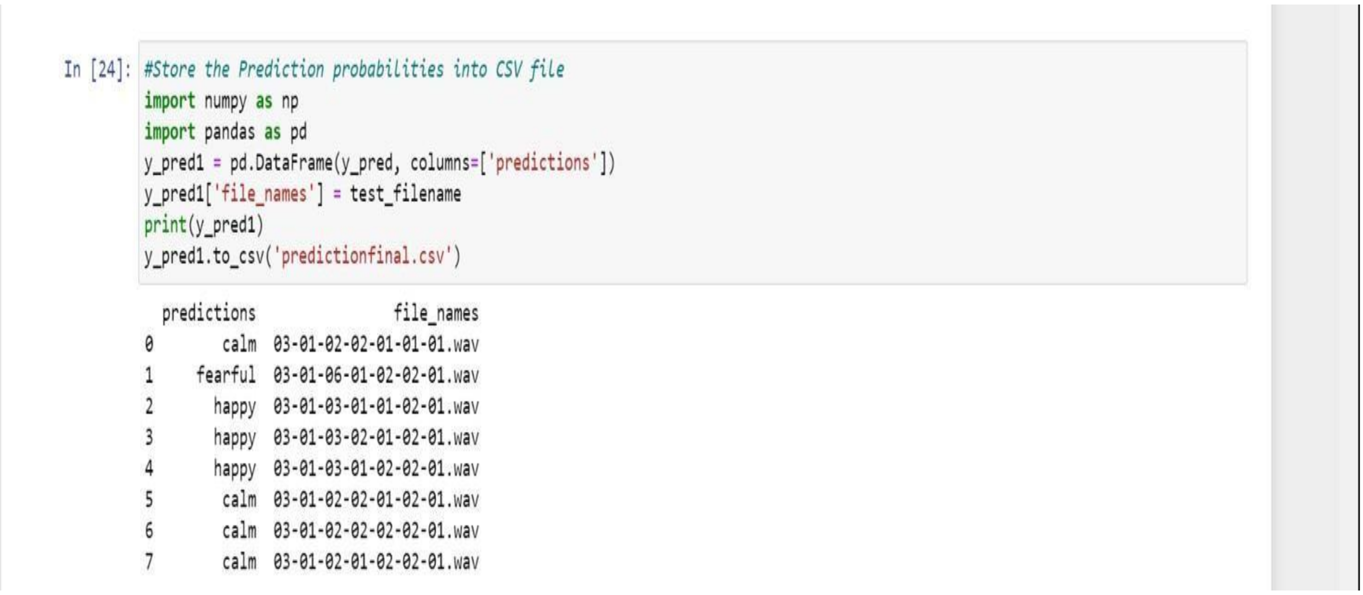
Step 10

Predicting the accuracy of trained model on dataset. Then we store the obtained result into the csv file.



```
In [30]: #predicting :
         y_pred=Emotion_Voice_Detection_Model.predict(x_test)
         y_pred

Out[30]: array(['calm', 'fearful', 'happy', 'happy', 'calm', 'calm', 'calm',
                'calm'], dtype='<U7')

In [31]: # DataFlair - Print the accuracy
         accuracy = accuracy_score(y_true=y_test, y_pred=y_pred)
         print("Accuracy: {:.2f}%".format(accuracy * 100))

         Accuracy: 87.50%
```

Fig.24 predicting the accuracy



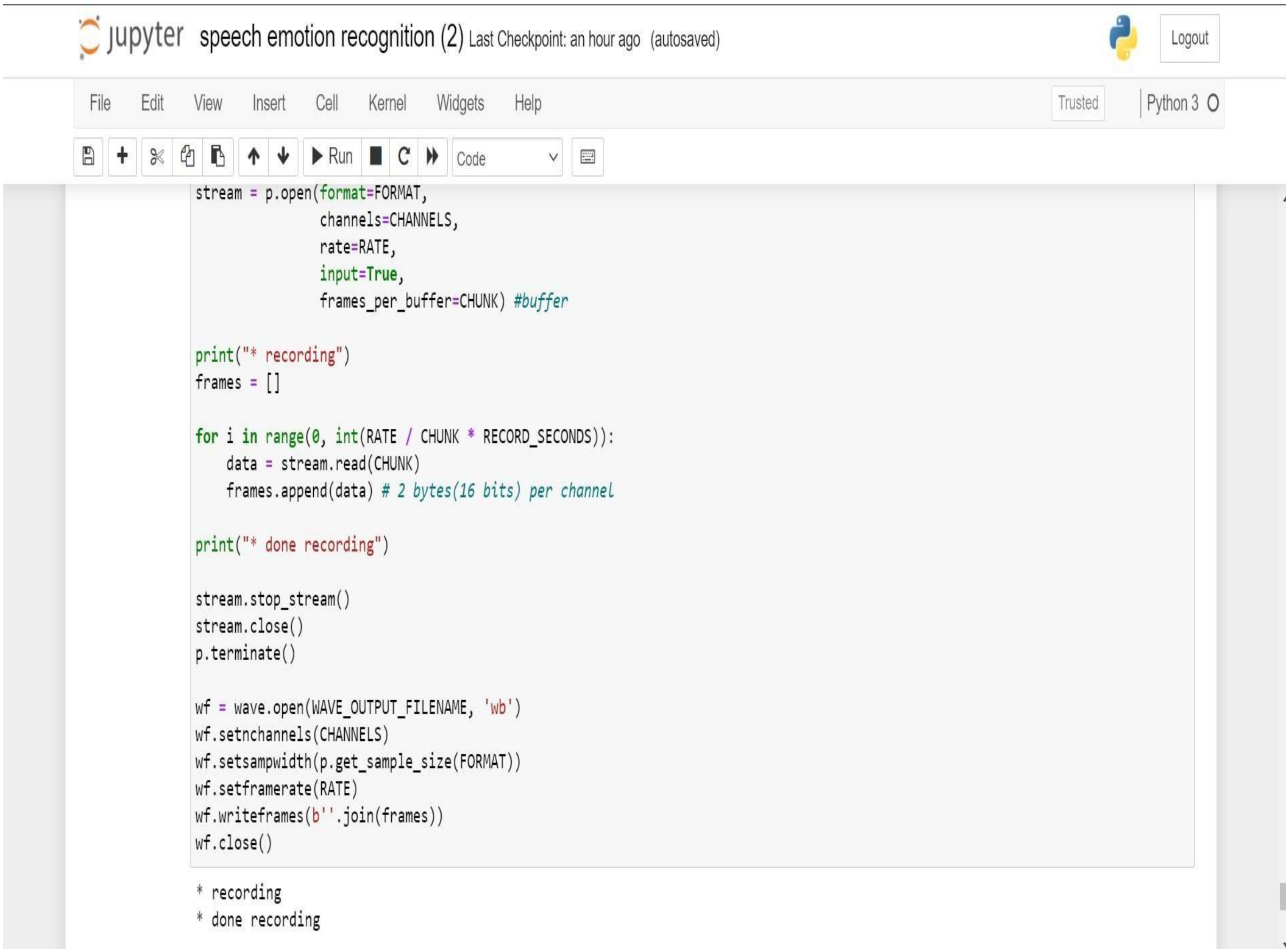
```
In [24]: #Store the Prediction probabilities into CSV file
         import numpy as np
         import pandas as pd
         y_pred1 = pd.DataFrame(y_pred, columns=['predictions'])
         y_pred1['file_names'] = test_filename
         print(y_pred1)
         y_pred1.to_csv('predictionfinal.csv')

           predictions                file_names
         0        calm  03-01-02-02-01-01-01.wav
         1     fearful  03-01-06-01-02-02-01.wav
         2       happy  03-01-03-01-01-02-01.wav
         3       happy  03-01-03-02-01-02-01.wav
         4       happy  03-01-03-01-02-02-01.wav
         5        calm  03-01-02-02-01-02-01.wav
         6        calm  03-01-02-02-02-02-01.wav
         7        calm  03-01-02-01-02-02-01.wav
```

Fig.25 Storing predictions into a csv file

4. working on a real-time audio input

Step 1

Recording the real-time audio of the user.



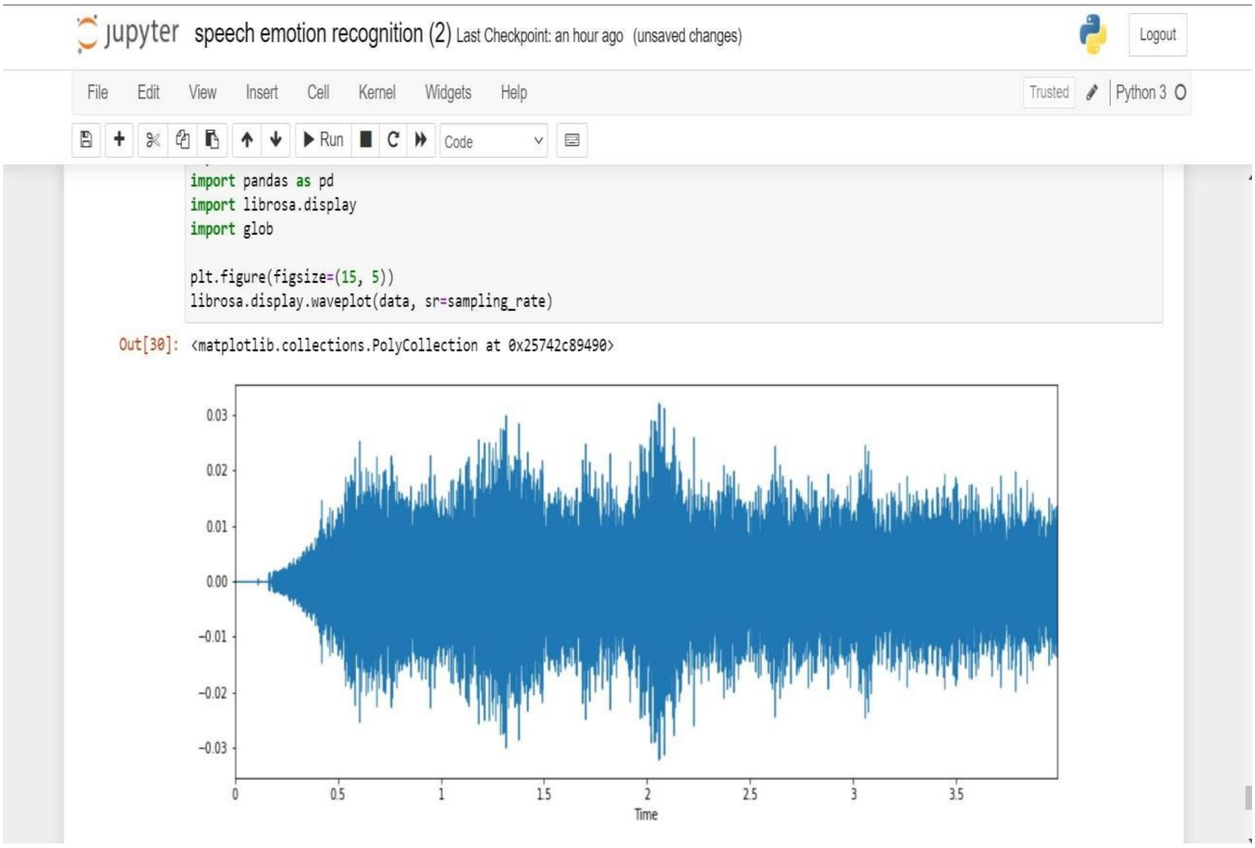
```
stream = p.open(format=FORMAT,
                channels=CHANNELS,
                rate=RATE,
                input=True,
                frames_per_buffer=CHUNK) #buffer

print("* recording")
frames = []

for i in range(0, int(RATE / CHUNK * RECORD_SECONDS)):
    data = stream.read(CHUNK)
    frames.append(data) # 2 bytes(16 bits) per channel

print("* done recording")

stream.stop_stream()
stream.close()
p.terminate()

wf = wave.open(WAVE_OUTPUT_FILENAME, 'wb')
wf.setnchannels(CHANNELS)
wf.setsampwidth(p.get_sample_size(FORMAT))
wf.setframerate(RATE)
wf.writeframes(b''.join(frames))
wf.close()
```

```
* recording
* done recording
```

Fig.26 Recording the audio of the user

Step 2

Plotting graphs to analyze the input that user provided.


jupyter speech emotion recognition (2) Last Checkpoint: an hour ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
0
0.5
1
1.5
2
2.5
3
3.5
Time

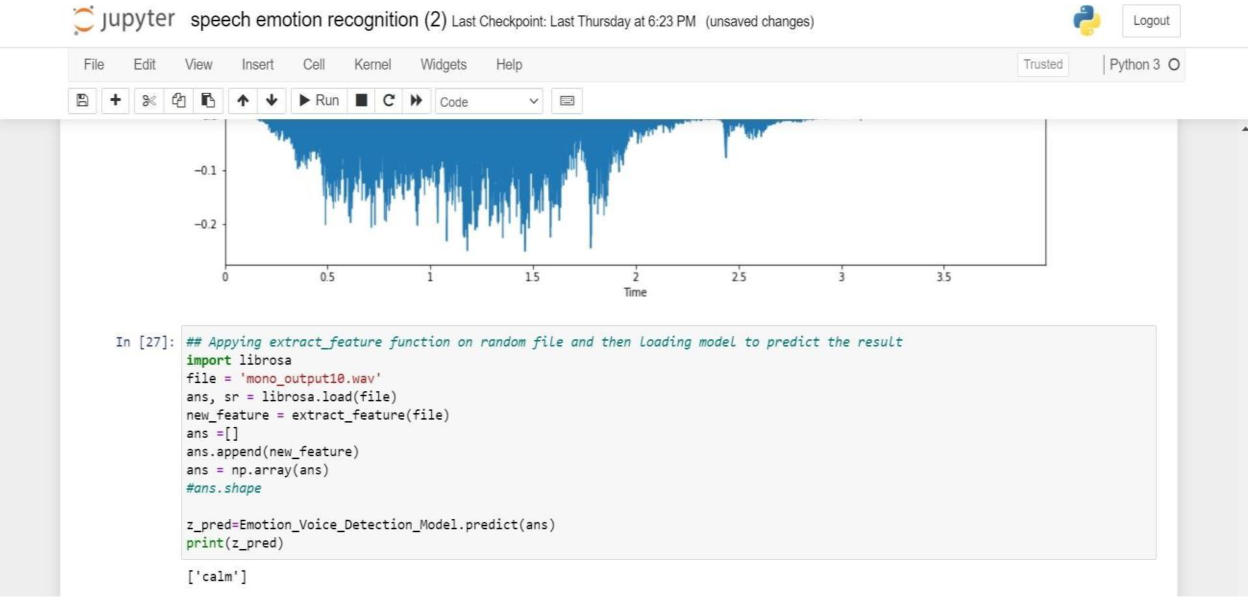
```
import pandas as pd
import librosa.display
import glob

plt.figure(figsize=(15, 5))
librosa.display.waveplot(data, sr=sampling_rate)
```

```
Out[30]: <matplotlib.collections.PolyCollection at 0x25742c89490>
```

Fig.27 Plotting graph of recorded audio

Step 3

Predicting the emotions of the audio provided.



```
In [27]: ## Appying extract_feature function on random file and then loading model to predict the result
         import librosa
         file = 'mono_output10.wav'
         ans, sr = librosa.load(file)
         new_feature = extract_feature(file)
         ans =[]
         ans.append(new_feature)
         ans = np.array(ans)
         #ans.shape

         z_pred=Emotion_Voice_Detection_Model.predict(ans)
         print(z_pred)

         ['calm']
```

Fig.28 graph predicting the emotions

ALGORITHM

- Load the input audios from the available database.
- pre-processing the dataset to remove unwanted data.
- Plotting graphs to analyze data.
- Training model with the help of MLP classifier.
- Predicting accuracy.
- Storing the predicted data in csv files.

CHAPTER 4

1. Poject Outcome:

- Project Report: Completed
- Research Paper: N/A
- Website: completed
- Patent: N/A

2. Importance

The key of expressing our feelings and emotions is communication. People use their body parts as well as their voice to communicate effectively. To portray one's feelings, gestures, body language, tone, and kindness are all employed simultaneously. Although, the oral form varies from one language to another, the non-verbal form of communication is the most common expression of common sense. Therefore, the technologies which are advanced are being developed to produce a natural environmental experience and includes speech understanding of emotional context. Advances in the realm of sensory processing have a positive impact on many applications. Performing the process of finding emotions include psychology, psychiatry, and neuroscience also benefits some research areas. These are the cognitive sciences departments depending on human interaction, in which the subject of learning is introduced by situations and questions, and based on the responses and answers, several are done.

2. Emotion difference between genders

A small but significant gender difference in emotional expression has been reported in adults, with women exhibiting greater emotional expression, especially positive emotions and including negative emotions such as sadness.Speech recognition is majorly based of different genders, of different age and races.This has a massive impact on peoples lives. These bias exist because of how we have

Structed our data, and our data analysis systems as likely to how cameras are program

Med to capture white skin tone, and even audio analysis strgulles with heavy voices

Causing problem with white male data.Potential relapses occur as few people are separated from presentations and are reluctant to communicate. Thus, replacement of traditional methods with computer- based acquisition program may benefit research. Similarly, the effective use of speech- based sensory processing is extensive. Assisted household items and helpers(Examples: Google Home and Amazon Alexa) are available in today’s era. Additionally, call centers supported by customer care often have automatic voice control, which may irritate the majority of their dissatisfied consumers. Rerouting these calls to a personal employee improves their customer service. Various programs include eLearning, online tutorial, survey, personal assistant (Example: Samsung S Voice and Apple Siri) etc. Themost recent application can be found in automatic or self-driving cars. They rely heavily on speech-activated controls. Unforeseen circumstances, such as nervousness, can induce a passenger to speak in hushed tones. Understanding the emotional content conveyed is crucial in these situations.Despite significant breakthroughs in artificial intelligence, it is extremely difficult to connect smoothly with robots in today's world, in part because machines they have zeroknowledge of feelings.

Speech-emotion recognition (SER), which seeks to extract emotions from voice signals, has recently been attracting increasing attention. This is a very complex task that elicits active emotional features as an open question. A feed-forward neural network, i.e., the Deep Neural Network (DNN) with more than one layer hidden between its inputs and outputs. It is able to learn a representation from of high-level non-flammable factors and effectively classify data. With adequate information of training and its various strategies, Deep Neural Networks perform well in various machine learning tasks (e.g., speech recognition). Feature analyzation in sensory perception is much less studied than in speech recognition. Many previous studies have identified aspects of emotional planning. In this study, DNN assumes the inclusion of common acoustic features in the speech part and generates a standardized level-level distribution, in which features of speech-level arebuilt and used to examine the level of speech quality. Since phase-level results already provide a lot of emotional detail and the classification of speech level does not involve much training, it is not necessary to use Deep Neural Networks for speech level separation. Instead, we use a neural network which is newly developed, called an advanced learning machine (ELM), to make emotional separation of speech level

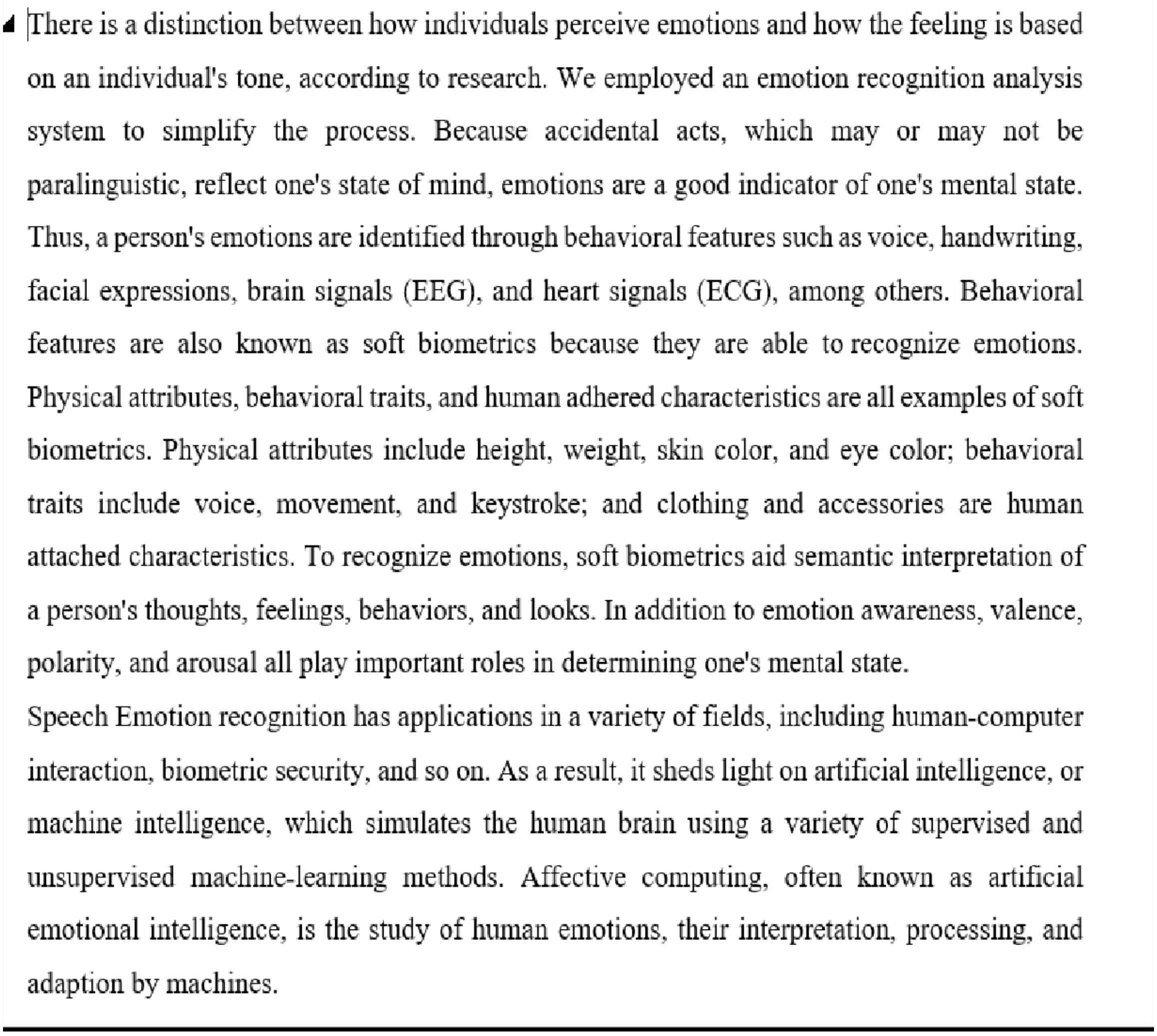
CHAPTER 5

CONCLUSION AND FUTURE SCOPE

Parameterization of audio-data for the purpose of emotion recognition. Various audio data were collected from several videos of human expressions gathered and turned into dataset. After this, conversion of data from audio to text files was done, various graphs were plotted between time and amplitude.

After this, in-depth graphs were plotted for this which leads towards the feature extraction. After this, another masking is performed for more accuracy.

FUTURE SCOPE:



There is a distinction between how individuals perceive emotions and how the feeling is based on an individual's tone, according to research. We employed an emotion recognition analysis system to simplify the process. Because accidental acts, which may or may not be paralinguistic, reflect one's state of mind, emotions are a good indicator of one's mental state. Thus, a person's emotions are identified through behavioral features such as voice, handwriting, facial expressions, brain signals (EEG), and heart signals (ECG), among others. Behavioral features are also known as soft biometrics because they are able to recognize emotions. Physical attributes, behavioral traits, and human adhered characteristics are all examples of soft biometrics. Physical attributes include height, weight, skin color, and eye color; behavioral traits include voice, movement, and keystroke; and clothing and accessories are human attached characteristics. To recognize emotions, soft biometrics aid semantic interpretation of a person's thoughts, feelings, behaviors, and looks. In addition to emotion awareness, valence, polarity, and arousal all play important roles in determining one's mental state.

Speech Emotion recognition has applications in a variety of fields, including human-computer interaction, biometric security, and so on. As a result, it sheds light on artificial intelligence, or machine intelligence, which simulates the human brain using a variety of supervised and unsupervised machine-learning methods. Affective computing, often known as artificial emotional intelligence, is the study of human emotions, their interpretation, processing, and adaption by machines.

5.1 Applications of the Major Project

- This project can be widely used in the field of healthcare.
- This project can also be widely used by any person who is having any difficulty with using technology
- Can be used in entertainment
- Can be used in high tech automobiles and smart phones
- It can be implemented in the field of education.

5.2 Future Work

- Note that although in our Project we have got a prediction accuracy, we are still researching on deep learning and neural network methods corresponding to this project as they can work better and more efficiently if the data set in hand was more complex and bigger.

- First, we are thinking of increasing the size of the current data set by a huge amount by gathering dataoff the internet or by putting some random manual data into it, then we will try to implement it for user input for massive amount of users in future.

- To build a working website with active user as well as handling ends.

References

[1] El.Ayadi, MS.Kamel, and F.Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases." *Pattern recognition* 44.3 (2011): 572-587.

[2] R.Khalil, E.Jones, MI.Babar,et al. "Speech emotion recognition using deep learning techniques: A review." *IEEE Access* 7 (2019): 117327-117345.

[3] Schuller, Björn, G.Rigoll, and Manfred Lang. "Hidden Markov model-based speech emotion recognition." *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).*. Vol. 2. Ieee, 2003.

[4] Nwe, T.Lay, S.Wei Foo, and L. C. De Silva. "Speech emotion recognition using hidden Markov models." *Speech communication* 41.4 (2003): 603-623.

[5] Huang, Zhengwei, et al. "Speech emotion recognition using CNN." *Proceedings of the 22nd ACM international conference on Multimedia*. 2014.

[6] Y. Dong, and L. Deng." *Automatic speech recognition".* Vol. 1. Berlin: Springer, 2016.

[7] H. Morris, and K. Stevens. "Speech recognition: A model and a program for research." *IRE transactions on information theory* 8.2 (1962): 155-159.

[8] Rabiner, Lawrence, and Bh. Juang. *Fundamentals of speech recognition*. Prentice-Hall, Inc., 1993.

[9] Nassif, A. Bou, et al. "Speech recognition using deep neural networks: A systematic review." *IEEE access* 7 (2019): 19143-19165.

[10] Graves, Alex, A.R. Mohamed, and G.Hinton. "Speech recognition with deep recurrent neural networks." *2013 IEEE international conference on acoustics, speech and signal processing*. Ieee, 2013.

[11] Han, Wei, et al. "Contextnet: Improving convolutional neural networks for automatic speech recognition with global context." *arXiv preprint arXiv:2005.03191* (2020).