ABSTRACT

We know that in today's world a lot of technologies train computers to think like a human mind . Artificial Intelligence or other activities contribute vitally in this development. That is what motivated us. We see a lot of chatbots are being made but wouldn't it be better to understand the emotion of the user . That is why we made a speech emotion recognition system that takes the sample speech input and responds to the emotion of the user based on parameters . Such a system can be utilized for various purposes where the emotions of the user play a major role. Special chatbots also can use this system.

CONTENTS

1. INTRODUCTION

1.1 PROBLEM DEFINITION

As human beings, speech is amongst the most natural ways to express ourselves. We depend so much on it that we recognize its importance when trying other communication forms like emails and text messages where we often use emojis to express the emotions associated with the messages. As emotions play a crucial role in communication, the detection and analysis of the same is of great importance in today's digital world of remote communication. Emotion detection is a significantly challenging task, because emotions are subjective. There is no common consensus on how to measure or categorize them. We define a SER system as a collection of methodologies that process and classify speech signals to detect emotions embedded in them. Such a system could be used in a wide variety of application areas like interactive voice based-assistant or caller-agent conversation analysis. In this study we attempt to detect underlying emotions in recorded speech by analysing the its features of the audio data of recordings.

1.2 GOALS AND OBJECTIVES

- This is an objective taken to understand human emotions and speech for the computers and bots for better analysis of human behaviour and responses. Understanding human speech emotion helps to respond in a much better way by analysing humans.
- Apply deep learning and using it to make a system that helps in speech emotion recognition
- To detect underlying emotions in recorded speech by analysing the features of the audio data of recordings

1.3 MOTIVATION AND SCOPE

The basic motivation in creating this project is the basic difference between humans and computers that computers cannot understand speech emotions and this provides a great scope for our project because of various help that our project would provide to the computers and bots to understand human speech emotions and respond in a much efficient way.

Through this project, we showed how we can use Machine learning to obtain the underlying emotion from audio data and some insights on the human expression of emotion through voice . This system can be employed in a variety of places like Call Centre for complaints or chatbots in linguistic research or marketing, in voice-based virtual assistants , etc.

2. LITERATURE SURVEY

1. **NAME OF THE PAPER**

Speech emotion recognition: Two decades in a nutshell benchmarks and ongoing trends", *Commun. ACM*, vol. 61, pp. 90-99, 2018.

B. W. Schuller,

INFORMATION - It tells in detail about speech emotion recognition trend

2. **NAME OF THE PAPER**

Emotion recognition using deep learning approach from audio–visual emotional big data", *Inf. Fusion*, vol. 49, pp. 69-78, Sep. 2019.

M. S. Hossain and G. Muhammad

INFORMATION - It tells about the use of audio and visual emotional big data analysis

3. **NAME OF THE PAPER**

Emotion communication system", *IEEE Access*, vol. 5, pp. 326-337, 2016.

M. Chen, P. Zhou and G. Fortino

INFORMATION - It tells about emotion communication analysis

3. ANALYSIS

3.1 MATHEMATICAL MODELING

MFCC- Mel-frequency cepstral coefficients

Mel scale

$$M(f) = 1125 \ln(1 + f/700) \qquad (1)$$

$$M^{-1}(m) = 700(\exp(m/1125) - 1) \qquad (2)$$

$$S_i(k) = \sum_{n=1}^{N} s_i(n)h(n)e^{-j2\pi kn/N} \qquad 1 \leq k \leq K$$

$$P_i(k) = \frac{1}{N}|S_i(k)|^2$$

Chroma

r^n=p

r= nth root pf p

where the ratio *r* divides the *p* (typically the octave, which is 2:1) into *n* equal parts.

3.2 REQUIREMENT ANALYSIS

3.2.1 FUNCTIONAL REQUIREMENT

The model should be able to do the following things . and on completion of so shows the implementation of project is done correctly

- It should be able to capture speech
- It should be able to analyse using deep learning
- It should understand the speech emotion using the model
- It should be able to correctly respond to the user with the speech emotion

3.2.2 NON-FUNCTIONAL REQUIREMENT

Here we specify some non-functional constraints that our project must follow

3.2.2.1 Performance

Response Time: The system shall give response in 1 second after posting the tweet

Capacity: The system must support 1000 users at a time.

User interface: The user interface screen shall response within 1 second

The system should be able to actively capture and collect samples provided by the user. The system needs to make an active response to the user as the late recognition reduces the user friendly nature of the system.

3.2.2.2 Security

1. User details -The system should keep all the user data safe .

3.2.2.3 Reliability

1. The user can use the application anytime he wants.
2. In case of unplanned system downtime, all features will be available again after one working day.

3.2.2.4 Availability

The system shall be available all the time.

3.2.2.5 Safety

. 1.The system should not detect any wrong emotions harming the user or raising any negative emotions

3.2.2.6 Software Quality

1. Good quality of framework = produces robust, bug free software which contains all the necessary requirements.
2. Customer satisfaction

3.2.2.7 Reusability

This application can be reused in Android & other platforms as a future scope.

3.2.2.8 Maintainability

The system shall provide the capability to back-up the data.

3.3 SDLC MODEL

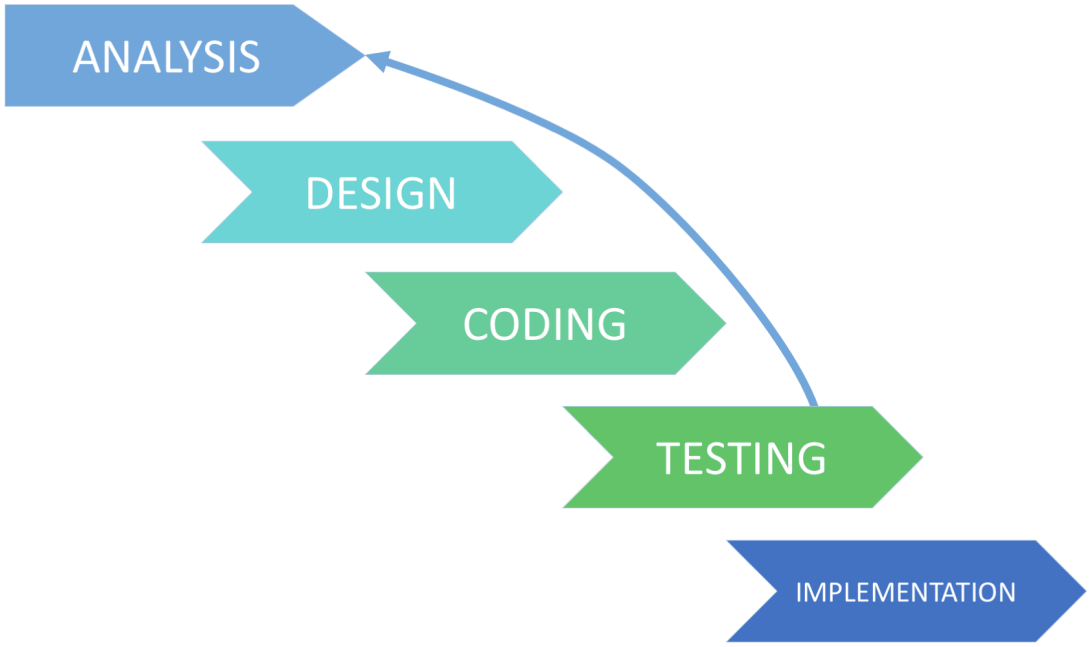

Figure 1 :- sdlc design

ANALYSIS :-

In the analysis phase we analysed the user requirement to understand upto what level this project needs to be implemented and gathering user needs was a big task . different data samples were collected.

DESIGN:-

In this phase GUI was made and all the designs were made to create something that is user friendly and looks attractive too. .

CODING:-

In this phase next all coding for the project was done for taking the speech sample extracting the features and emotion selection. In this part the coding language was also chosen to make the coding part easy.

TESTING:-

In the testing part various modules were tested where extracted features were checked and also various samples were tested and checked whether the emotion given is correct or not.

IMPLEMENTATION:-

Finally all modules were combined and the project was completely ready after the testing phase .

4. SOFTWARE REQUIREMENT SPECIFICATION

4.1 Software Requirements

The project utilizes following software requirements

1. **MLP algorithm**

   A multilayer perceptron (MLP) is a class of feedforward artificial neural network (ANN). The term MLP is used ambiguously, sometimes loosely to *any* feedforward ANN, sometimes it refers to multiple layers networks of perceptrons (with threshold activation); see § Terminology. Multilayer perceptrons are sometimes colloquially referred to as "vanilla" neural networks, especially when they have a single hidden layer.

2. **User interface anvil**

   Anvil fills in the gaps by allowing you to build a full-stack web app using only Python. You can build a UI with a simple drag and drop UI (or build it with code if you insist), plot with your favorite Python library for plotting (Plotly, Matplotlib, etc.), and then deploy to the web in immediately. No servers or containers to deal with.

   Often a large part of your app will be sharing insights through data visualization. Anvil supports plotting from nearly every popular Python plotting library. Here are the libraries that Anvil has made guides for

   - Matplotlib
   - Plotly
   - Seaborn
   - Bokeh
   - Altair
   - Pygal

3. **Python**

   Python is a high-level general-purpose interpreted programming language. Python's design emphasizes on code readability with its notable use of significant indentation. Its

object-oriented approach as well as language constructs aim to help programmers write clear, logical code for small and large-scale projects.

Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including object-oriented structured (particularly, procedural), and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.

4. **Dataset**

   For this Python mini project, we'll use the RAVDESS dataset; this is the Ryerson Audio-Visual Database of Emotional Speech and Song dataset, and is free to download. This dataset has 7356 files rated by 247 individuals 10 times on emotional validity, intensity, and genuineness. The entire dataset is 24.8GB from 24 actors, but we've lowered the sample rate on all the files

4.2 Hardware Requirements

- CPU (to control the use interface)
- Keyboard and Mouse (for user input)
- Display (used by customer for performing the transactions)
- Operating System: Windows
- Hard disk: 4GB RAM: 2GB
- Processor: Pentium Dual Core CPU

5. SYSTEM DESIGNS

5.1 SYSTEM ARCHITECTURE

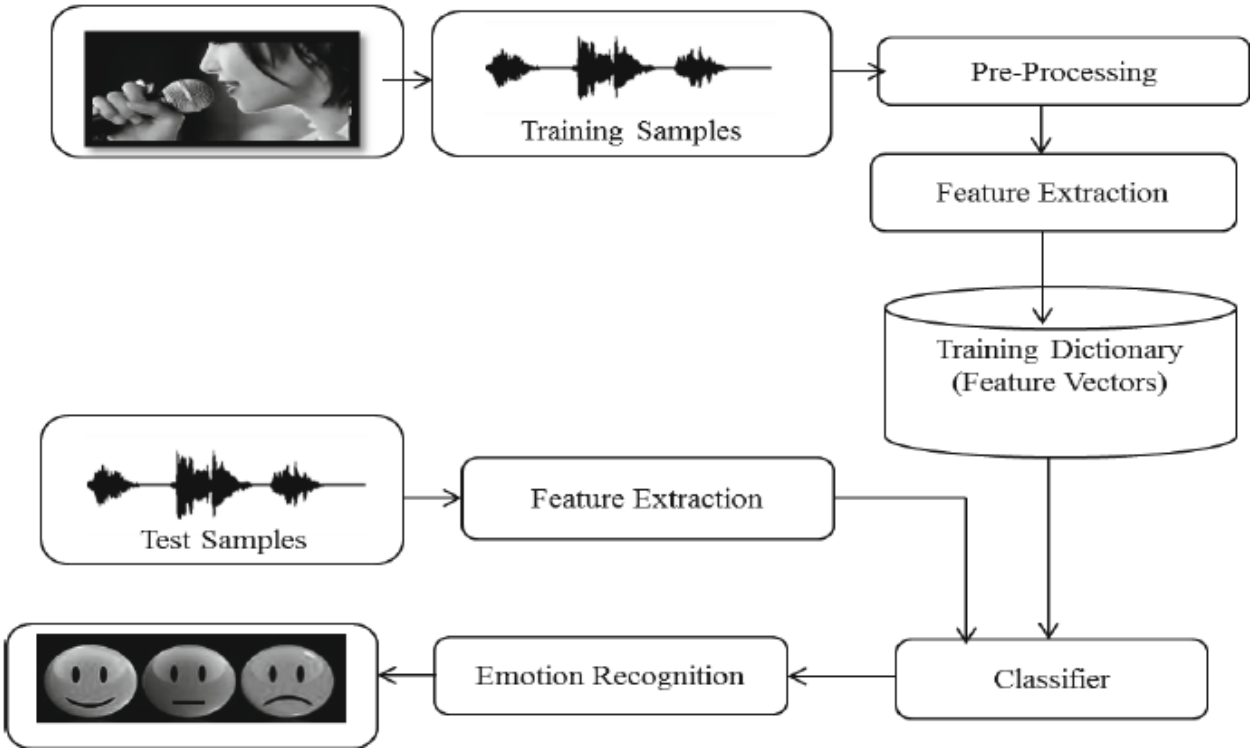

Figure 2 - system architecture

In the system architecture the overall system shows that the model is firstly trained by using the dataset of voice samples then preprocessing is done then feature extraction like mfcc , mel , chroma is done and then they are passed to the classifier then it is tested with test voice samples and then it features are extracted and emotion like calm, happy, fearful, disgust is recognized from it and displayed to user on the webapp for that we are using anvil for taking the input and displaying the output.

5.2 UML DIAGRAMS

5.2.1 STATE DIAGRAM

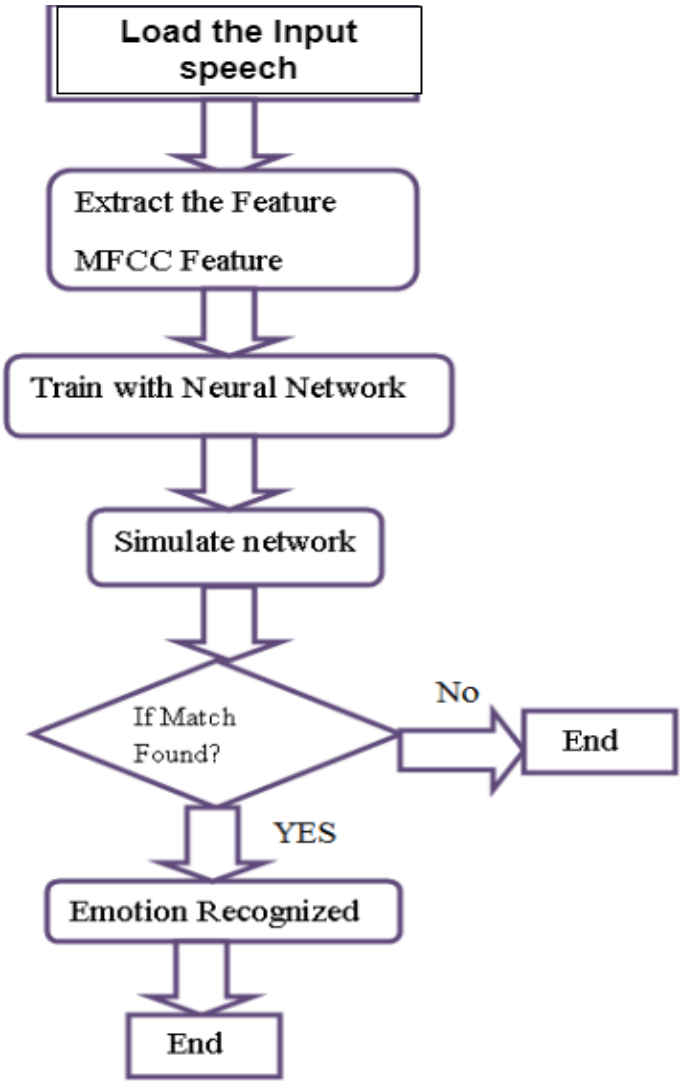

Figure 3-state diagram

The state diagram depicts various states involved in the project first a voice sample is taken . Features like mfcc, mel, chroma are used to train the model using the neural network , backpropagation is used to adjust weights to get correct output then if a match is found the emotion is recognized.

5.3 USE CASE DIAGRAM

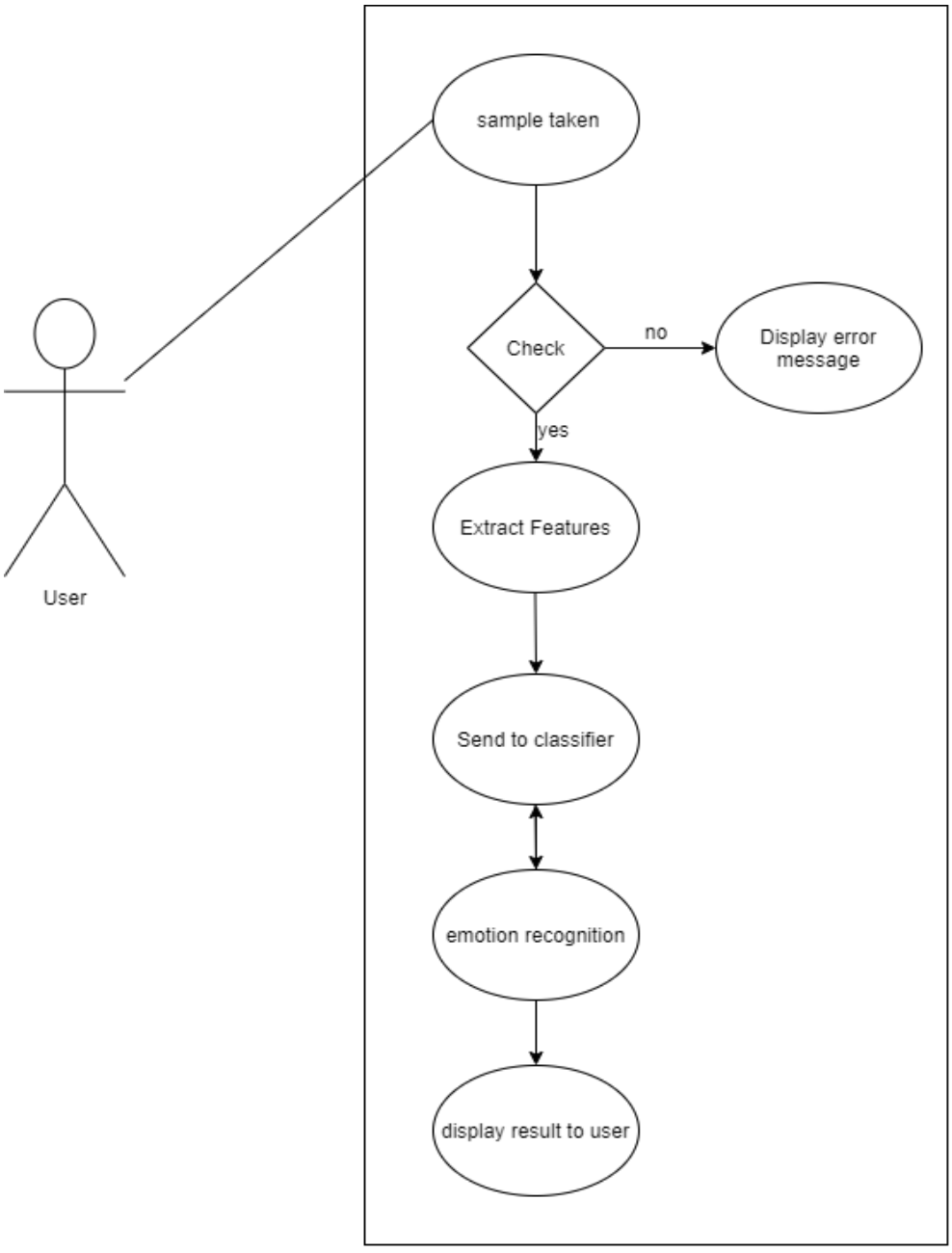

Figure 4:- Use case Diagram

The use case diagram shows how the user is using the system. User clicks on the site on the select sample option. Selects a voice sample. Then a voice sample is uploaded then its features are extracted and sent to the classifier then emotion is recognized and the result is displayed.

6. MINI PROJECT IMPLEMENTATION

6.1 OVERVIEW OF PROJECT MODULES

The project uses the following modules for the system

1. Speech capture module:- this module captures or uses existing speech samples and which are further sent to next module for extraction of its features

2. Feature extraction module:- this module takes the speech sample and uses it to extract features mathematically and logically like mfcc and others.

3. Emotion recognition module:- this module from the features extracted analyses the existing data and emotions set and returns the emotion to the user.

6.2 TOOLS AND TECHNOLOGIES

The following tools and technologies are used for creation of this project:-

1. MLP
2. Python
3. Google colab

6.3 ALGORITHMIC DETAILS

A multilayer perceptron (MLP) is a class of feedforward artificial neural network (ANN). The term MLP is used ambiguously, sometimes loosely to *any* feedforward ANN, sometimes it refers to multiple layers networks of perceptrons (with threshold activation); see § Terminology. Multilayer perceptrons are sometimes colloquially referred to as "vanilla" neural networks, especially when they have a single hidden layer.

The term "multilayer perceptron" does not mean a single perceptron that has multiple layers. Rather, it means many perceptrons that are organized into layers.

An alternative is "multilayer perceptron network". Moreover, MLP "perceptrons" are not perceptrons in most cases.

True perceptrons that use a threshold activation function such as the Heaviside step function. MLP perceptrons can employ arbitrary activation functions.

A true perceptron performs *binary* classification, an MLP neuron is free to either perform classification or regression, depending upon its activation function.

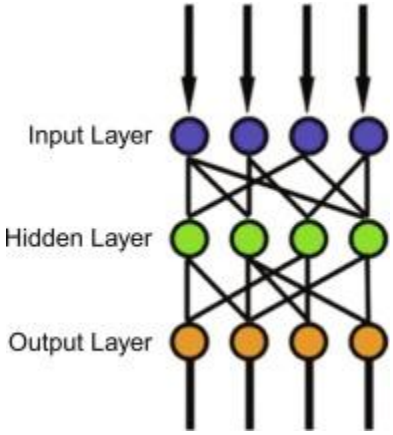

Figure 5:- MLP algorithm

7. SOFTWARE TESTING

Software Testing is a method to check whether the software product matches expected requirements and to make sure completely that the software product is Defect free. It involves running of software/system components using automated or manual tools to evaluate errors in the system. The use of software testing is to identify gaps , errors or missing requirements in contrast to actual requirements.

Software Testing is Important because to find any bugs and solve them early. Tested software products ensure security ,reliability and high performance which further results in time saving, cost effectiveness and customer satisfaction.

Test cases are as follows :

TEST CASE SCENARIOS	GIVEN INPUT	EXPECTED OUTPUT	ACTUAL OUTPUT	COMMENT
An input speech sample with emotion anger was given	Speech sample	emotion-anger	emotion-anger	correct
An input speech sample with emotion calm was given	Speech sample	emotion-calm	emotion-calm	correct
An input speech sample was given	Speech sample	emotion-	emotion-	correct

8. RESULTS

8.1 OUTCOMES

The outcome of this project is a system that takes speech samples as input and extracts features from it like mfcc and based on such features it analyses and returns the emotion of the speech sample provided.

8.2 SCREENSHOTS

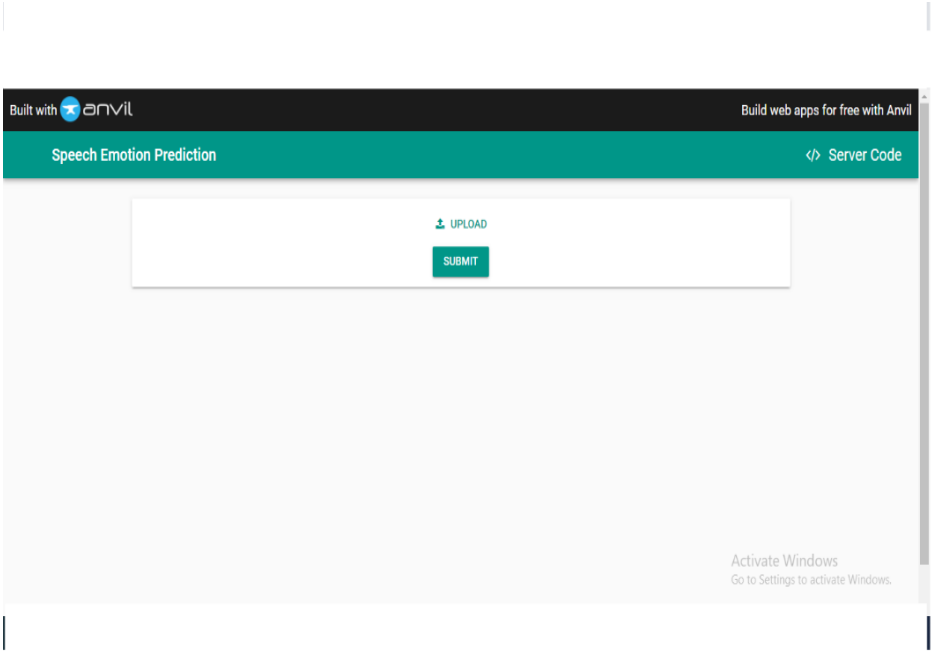

Figure 6 :-output1.

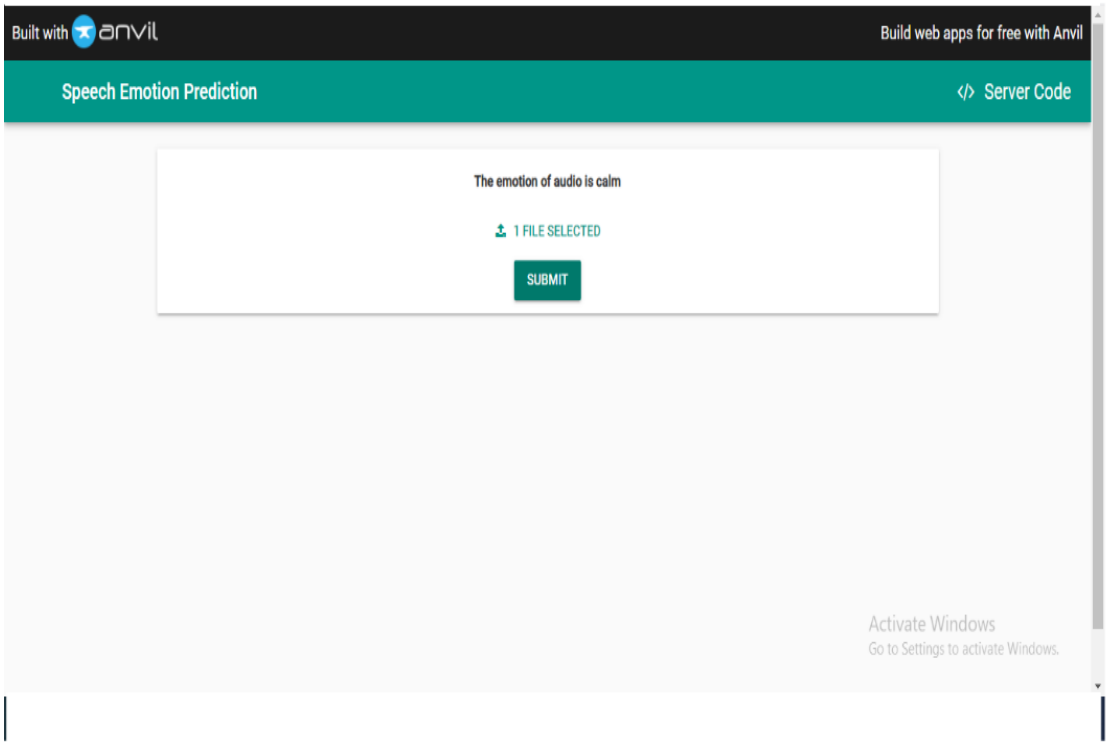

Figure 7 :-output 2.

9. CONCLUSION

9.1 CONCLUSION

Finally we have concluded that a system like speech emotion recognition can be made with taking a speech sample and extracting features from it and based on the given set of values for the features an emotion can be returned as output .

9.2 FUTURE WORKS

- .The future scope of this project is to increase the efficiency of the project and reduce the delay time for output.
- .An accurate implementation of the pace of the speaking can be explored to check if it can resolve some of the deficiencies of the model.
- .Figuring out ways to clear random silence from the audio clip.
- .Exploring other acoustic features of sound to check their applicability in the domain of speech emotion recognition. These features could simply be some proposed extensions of MFCC like RAS-MFCC or they could be other features entirely like LPCC, PLP or Harmonic cepstrum.
- .Following lexical feature based approach towards SER and using an ensemble of the lexical and acoustic models. This will improve the accuracy of the system because in some cases the expression of emotion is contextual rather than vocal.
- .Adding more data volume either by other augmentation techniques like time-shifting or speeding up/slowing down the audio or simply finding more annotated audio clips.

10.REFERENCES

1.Speech emotion recognition: Two decades in a nutshell benchmarks and ongoing trends", *Commun. ACM*, vol. 61, pp. 90-99, 2018.

2.Emotion recognition using deep learning approach from audio–visual emotional big data", *Inf. Fusion*, vol. 49, pp. 69-78, Sep. 2019.

3.Emotion communication system", *IEEE Access*, vol. 5, pp. 326-337, 2016.

4.https://www.analyticsvidhya.com/blog/2020/12/mlp-multilayer-perceptron-simple-overview

5.https://machinelearningmastery.com/when-to-use-mlp-cnn-and-rnn-neural-networks

11. Plagiarism Report

Report Title:	BE_SCOA_Project_Plagerism_Report
Report Link: (Use this link to send report to anyone)	https://www.check-plagiarism.com/plag-report/320692c0c3ca4365087e65ff94f3f54ac08ac1624021058
Report Generated Date:	18 June, 2021
Total Words:	3649
Total Characters:	19368
Keywords/Total Words Ratio:	0%
Excluded URL:	No
Unique:	91%
Matched:	9%

Fig 8:Plagiarism Report