



INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
SONEPAT

Speech Emotion Recognition


SUPERVISED BY-
Dr. Jitendra Samriya

SUBMITTED BY-
Gaurav - 12111073
Abhay Pratap Singh - 12111089
Adarsh Chaudhary - 12111093





Objectives

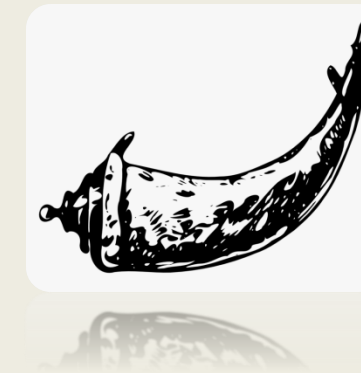
- To build a model to recognize emotion from speech using librosa and sklearn libraries and the RAVDESS dataset.
 - Detect emotions of speech and tells the accuracy of our model based on the dataset provided.
 - To deploy the model on web application using flask.
- 

Introduction

- Human machine interaction is widely used nowadays in many applications. One of the medium of interaction is speech. The main challenges in human machine interaction is detection of emotion from speech.
- Emotion can play an important role in decision making. We propose a system that will help to analyze different speech signals and gather the emotions from the same efficient solution based on various combinations.
- An emotional speech RAVDESS dataset is selected then emotion specific features are extracted from those speeches and finally a MLP classification model is used to recognize the emotions.

Tools Used:

- Python
- JupyterLab
- Librosa
- Sklearn
- MLP classifier/Tensorflow
- Flask.



Speech Emotion Recognition

Speech Emotion Recognition, abbreviated as SER, is the act of attempting to recognize human emotion and affective states from speech. This is capitalizing on the fact that voice often reflects underlying emotion through tone and pitch.

We define an SER system as ‘ ***a collection of methodologies that process and classify speech signals to detect emotions embedded in them*** ’. Such a system can find use in a wide variety of application areas like interactive voice-based-assistant or caller-agent conversation analysis.

Methodology

 **Preprocessing** **Feature Extraction** **Classification**

1. Pre-processing:

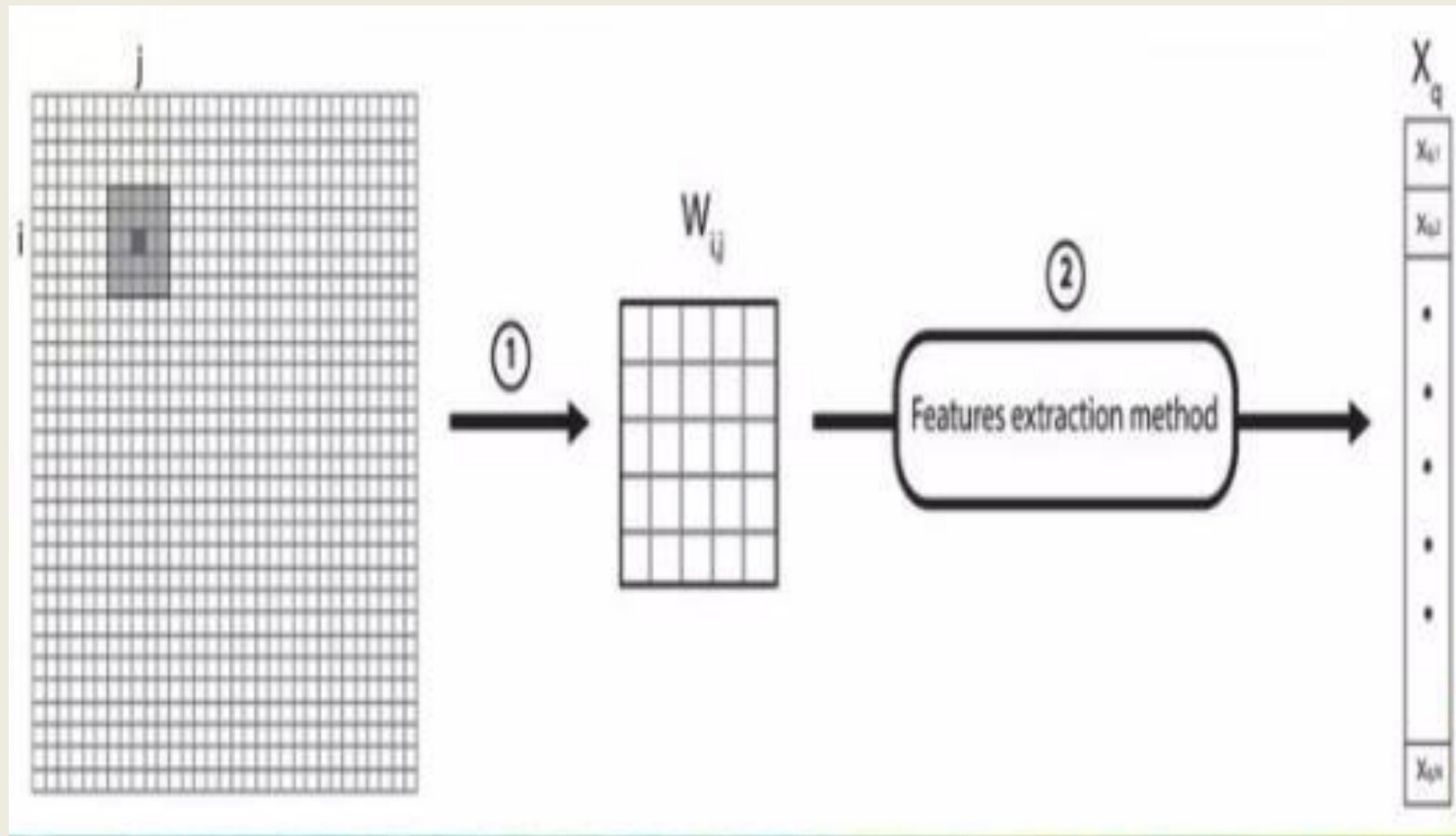
The removal of unwanted noise signal from the speech.



- **Silent Removal**
- **Background Noise Removal**
- **Windowing**
- **Normalization**

2. Feature Extraction

- Extract the feature from audio file.
- Used to identify how to speak.



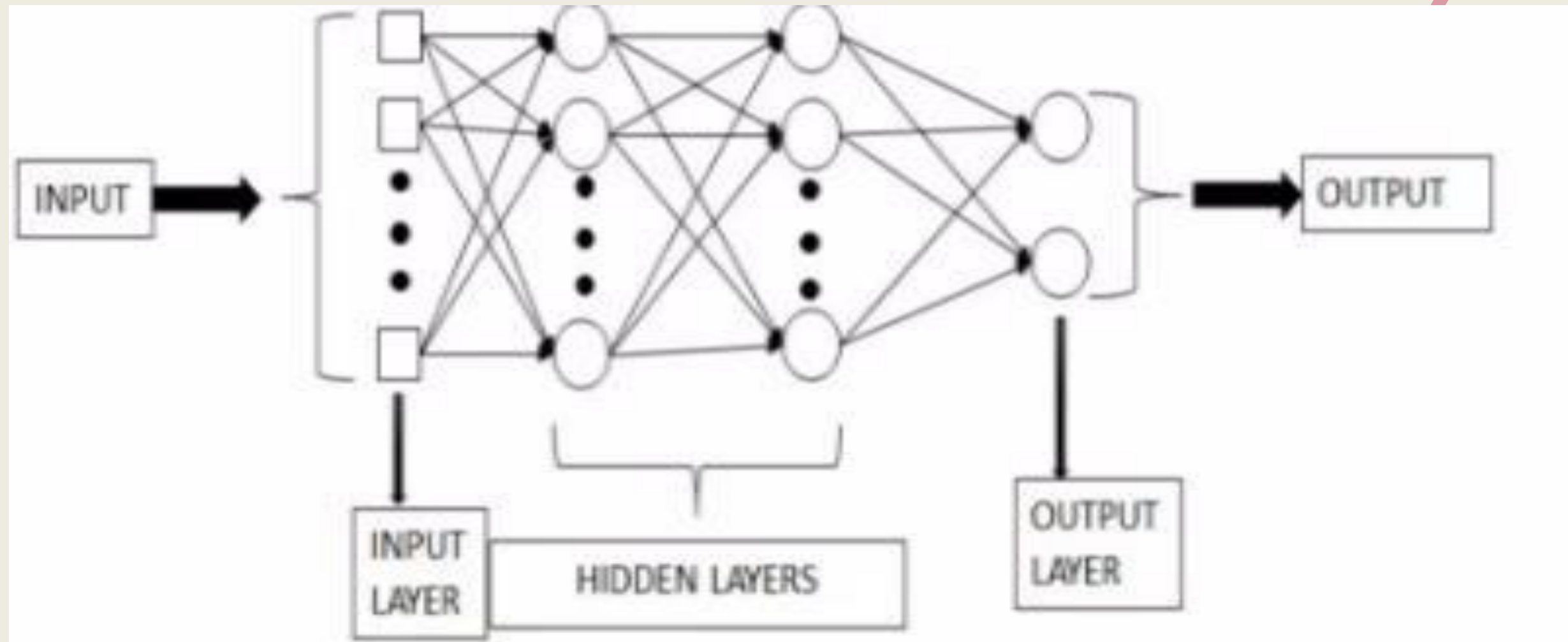
- Pitch
- Loudness
- Rhythm etc.

3. Classification

The third step is the main step of the system in which the audio speech is classified into different emotions based on the features extracted from the audio speech. With the help of the features extracted, the audio speech is classified into different emotions.

In our project we will be using MLP classifier.

- Match the feature with corresponding emotions.



Multi-Layer Perceptron Classifier

- A **multilayer perceptron** (MLP) is a class of feedforward artificial neural network (ANN).
- MLP consists of at least three layers of nodes-input layer, hidden layer and output layer.
- MLPs are suitable for classification prediction problems where inputs are assigned a class or label.

Building the MLP Classifier involves the following steps-

1. Initialization of MLP Classifier.
2. Neural Network.
3. Prediction.
4. Accuracy Calculation.

❑ DEPLOYMENT OF MODEL USING FLASK

Flask is a web micro-framework used to create web services that expose machine learning model as APIs.

Why Flask?

Easy to use.

Built in development server and debugger.

Integrated unit testing support.

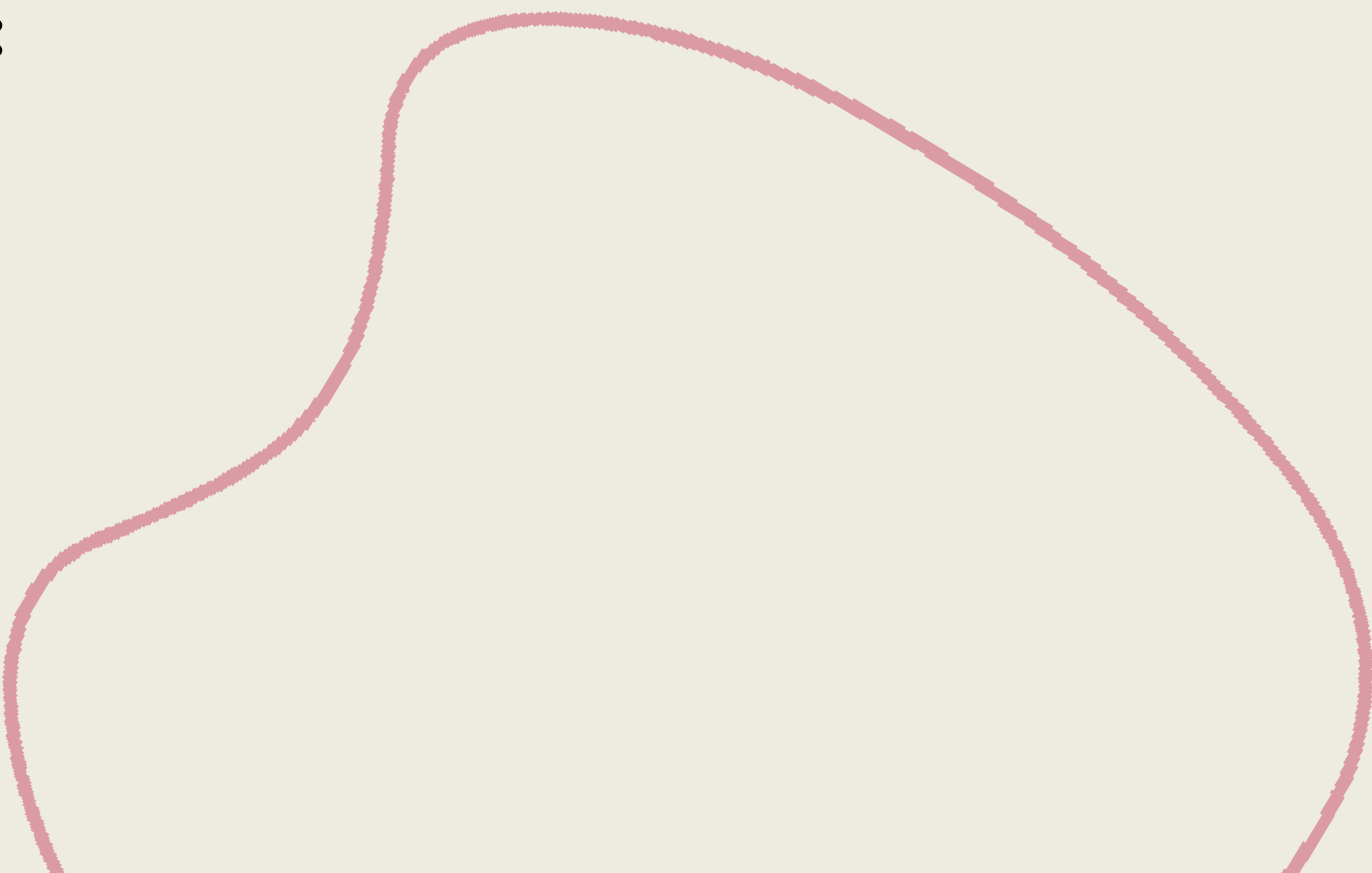


Flask

Deployment on web

Deploying a speech emotion and recognition model on the web involves creating a web application that allows users to interact with the model through a web interface.

Steps to deploy model on web:

1. Prepare the trained model.
 2. Build a web application.
 3. Create an input form.
 4. Process the input data.
 5. Make predictions using the model.
 6. Return the prediction to the user.
 7. Deploy the web application.
- 

Limitations

- Different set of words for each language make the process of training different for the language. Sometime meaning of the words also changed with place in the sentence.
- The emotion recognition system is subjective in nature. It means that the system works for one language may not work well for another language.

Conclusion

In this model we proposed a system that will analyse the speech signal and identify the emotion present in the speech or signal using concepts of Machine Learning where we will use Python as a language and RAVDESS as a data set.

- This model will analyse the audio file and return us the efficiency of the system.



Thank you!