Machine Learning (CS-584)

*Project Final Report*

# NAME: SASHANK LAKSHMANA BOMMADEVARA

# CWID: A20491471

### PROJECT TITLE:

*SPEECH EMOTION RECOGNITION With Librosa*

*OBJECTIVE*

*To build a model to recognize emotion from speech using the librosa and sklearn libraries and the RAVDESS dataset.*

# **MOTIVATION**

Have you ever made use of a phone service? What did you think of the experience? It may be really frustrating to have a robot ask you questions. Assume you are dissatisfied and call the company again, only to be met with a robot on the other end of the line. That is an example of how machine learning could be used to understand speech emotion and improve customer service. Emotional programming for robots has been identified as a critical step in making machines appear and behave like humans.

### **INTRODUCTION**

The technique of attempting to recognize affective and emotional states in human speech is known as speech emotion recognition. This takes use of the fact that voice tone and pitch usually reveal underlying emotion. Animals such as dogs and horses use this phenomenon to understand human emotion.

Emotion recognition is becoming more important as user experience and the use of Voice User Interfaces increase (VUIs). There are practical uses for developing emotion recognition systems based on speech. When the system is employed in real-world applications, however, the real-world background noise reduces the efficacy of speech-based emotion recognition.

SER (Speech Emotion Recognition) is used in contact centers to classify conversations based on emotions. It may also be used as a performance measure for conversational analysis, detecting dissatisfied customers, customer satisfaction, and so on.

It may also be employed in-car board systems based on information supplied by the driver's mental state to the system to begin his/her safety, avoiding accidents from occurring.

Telemedical application for determining the caller's mental condition.

# What is Librosa?

A Python library for analyzing audio and music is called librosa. It has backwards compatibility, a flatter package layout, standardized interfaces and names, modular functions, and readable code.

# What is Sklearn?

Python's Scikit-learn (Sklearn) library for machine learning is the most effective and reliable. Through a consistent Python interface, it offers a variety of effective techniques for statistical modeling and machine learning, such as dimensionality reduction, clustering, and classification. NumPy, SciPy, and Matplotlib serve as the foundation for this library, which was primarily constructed in Python.

# ***DATASET***

For this Machine Learning project, we'll use the RAVDESS dataset, which stands for Ryerson Audio-Visual Database of Emotional Speech and Song and is accessible for free download. 247 participants rated 7356 files in this dataset ten times for emotional validity, intensity, and sincerity. The entire dataset, which includes 24 actors, is 24.8GB in size, but we've lowered the sampling rate across the board.

# ***DATA ANALYSIS AND PROJECT EXECUTION***

The whole pipeline is as follows:

Preparing the Dataset: In this step, we download and transform the dataset so that it is ready for extraction.

Loading the Dataset: This procedure comprises loading the dataset in Python and collecting audio characteristics from the voice signal, such as power, pitch, and vocal tract configuration. We will use the librosa module to do so.

After we have prepared and loaded the dataset, we simply train it on a suitable sklearn model.

Model Validation: Measuring the performance of our model.

Extractable Features in Audio Signals used to Analyze Audio Signals

MFCC: Mel Frequency Cepstral Coefficient, represents the short-term power spectrum of a sound.

Chroma: Chroma is a method of converting sounds into numerical numbers. Most of the time, Chroma can be converted to a vector data format. A overview of Chroma history comprises the feature extraction process and can become an important aspect of data engineering. Chroma is a 12-element vector that calculates energy based on sound pitch.

STFT: Short-Time Fourier Transform

Mel: Spectrogram Frequency

# MACHINE LEARNING CLASSIFIER USED: MLPCLASSIFIER

Artificial Neural Networks, or ANNs, are widely used in many applications today, including classification, and there are many libraries and frameworks dedicated to building Neural Networks with ease. Most of these frameworks and tools, however, require many lines of code to implement when compared to the simple Scikit-Learn library that we will learn now.

Other Classification Algorithms vs. MLPClassifier

MLPClassifier is an abbreviation for Multi-layer Perceptron Classifier, which links to a Neural Network. Unlike other classification methods such as Support Vectors or Naive Bayes Classifier, MLPClassifier does classification using an underlying Neural Network.

However, one similarity to Scikit-other Learn's classification algorithms is that implementing MLPClassifier requires no more work than implementing Support Vectors, Naive Bayes, or any other Scikit-Learn classifier.

Perceptron (neurons) are layered in numerous layers in an MLP. Every node on each layer is linked to every other node on the following layer. A single layer has no connections between nodes. A fully (densely) connected neural network is referred to as an MLP (FCNN). To add layers, we utilize Keras' Dense() class. Data flows from the input to the output of an MLP in one (ahead) direction across layers. In certain publications, an MLP is also referred to as a Feed-Forward Neural Network (FFNN) or a Deep Feed Forward Network (DFFN). A sequential model is a form of MLP. To create MLPs, we utilize Keras' Sequential() class.

# ***Getting predictions***

After Splitting the Data Using Train-Test-Spilt with ratio 80:20 and Initializing the MLP Classifier. The MLP Classifier is trained and used to predict for test dataset

The Accuracy Observed is 75%

# ***CONCULSION***

Hence, the MLP Classifier achieved good accuracy and satisfied in prediction of output with minimal loss. So, I conclude that the MLP Classifier is one of the speech recognitions models which can yield good results.

# ***FUTURE WORK***

To Build a model that can identify emotion from Natural speech instead of professional environment with minimal noise and environmental interference

# ***THE END***