Speech Emotion Recognition Using MLP Classifier

By:

Divesh Gadhvi

Parth Borkar

Tanay

Koutilya Pande

Abstract

Speech is the most fundamental means of communication for humans, and language is their primary form of expression. In social connection, emotions are very important.

Introduction

The importance of human-computer interactions (HCI) cannot be overstated in the rapidly evolving AI world. Siri and Alexa are physically closer to us than other people in our world. Physical and virtual service robots will soon be more prevalent in society, doing tasks like judging the success of your marketing campaign and caring for the elderly. Having a deeper awareness of human emotions lays the way for better service delivery and a greater understanding of people's needs.

Human emotions are simple for us to understand since we are normal people, but they are more challenging for machines to comprehend. We use machine learning to train a machine how to recognize emotions so that it can comprehend them. An algorithm that creates a model using training data and then uses it to make decisions or predictions without having to be told to. Emotion Detection from voice aims to identify the emotions of a particular person from an input voice sample.

Processes involved in emotion detection include feature extraction from audio data, followed by the application of different classifiers. Our emotions include a variety of acoustical characteristics, including Energy, Pitch, Rhythm, and Loudness. By removing undesirable noise signals from these audio files using a variety of pre-processing techniques known as feature extraction, we must extract all these characteristics from the provided audio file. By utilizing the features of the librosa module, we can convert our audio files into digital data. For example, the Mel feature will be used to capture characteristics of the frequency of the signal represented on the Mel scale, the Mfcc feature will be used to describe the spectrum of the spectrum, or the short-term power spectrum, of the given input audio file, and the Chroma feature will be used to capture melodic and harmonic characteristics of sound based on pitch of the given input. We may use a variety of classifiers to match these features with the corresponding emotions after extracting all these features.